# Reference Magnetism Beyond the Predicate: Two Putnam-Style Results

Rohan Sud†

Many contemporary philosophers accept David Lewis's claim that, among the candidate meanings for our predicates, some are more natural than others – they do better or worse at "carving nature at its joints".[1] Call this claim *predicate naturalism*. However, among those that accept predicate naturalism, disagreement remains over a further question: does the notion of naturalness extend "beyond the predicate"?[2] That is, just as the candidate meanings of predicates can be more or less natural, can the candidate meanings of logical vocabulary also be more or less natural? Call the affirmative answer *logical naturalism*. Several authors – most prominent of which is Ted Sider – have argued in support of logical naturalism. And they have won many converts for their efforts.[3] But many others remain skeptical.

Predicate naturalism, it is thought, helps rebut various radical indeterminacy arguments associated with Hilary Putnam and Saul Kripke's reading of Wittgenstein.[4] It does so in combination with a popular meta-semantic theory, originally considered by David Lewis, called *reference magnetism*.[5]

This has been an influential motivation for predicate naturalism. My aim in this paper is to show that the same threats of radical indeterminacy rearise for proponents of reference magnetism and predicate naturalism – threats which logical naturalism rebuts. In other words, I'll argue that, in so far as we are moved to accept predicate naturalism

†Forthcoming in *Journal of Philosophy*. Accepted September 2022. Thanks to Maegan Fairchild, Boris Hennig, Mark McCullah, and Bradley Richards for helpful discussion and comments. Thanks also to two anonymous referees. Special thanks to Dan Hoek who provided invaluable advice on streamlining and simplifying the proofs in the appendices.

[1] David Lewis, "New Work for a Theory of Universals," *Australasian Journal of Philosophy*, LXI, 4 (December 1983): 343–77.

[2] The question is raised and pursued in Theodore Sider, *Writing the Book of the World* (Oxford: Oxford University Press, 2011).

[3] *Ibid.*; and Theodore Sider, "Ontological Realism," in David Chalmers, David Manley, Ryan Wasserman, eds., *Metametaphysics* (Oxford: Oxford University Press, 2009), pp. 384–423.

[4] Hilary Putnam, "Models and Reality," *The Journal of Symbolic Logic*, XLV, 3 (September 1980): 464–82; Hilary Putnam, "Realism and Reason," *Proceedings of the American Philosophical Association*, L, 6 (August 1977): 483–98; Hilary Putnam, *Reason, Truth and History* (Cambridge: Cambridge University Press, 1981); Saul Kripke, *Wittgenstein on Rules and Private Language* (Oxford: Blackwell, 1982).

[5] Lewis, "New Work for a Theory of Universals," *op. cit.*; David Lewis, "Putnam's Paradox," *Australasian Journal of Philosophy*, LXII, 3 (September 1984): 221–36.

by its ability to fend off threats of radical indeterminacy, we should also accept logical naturalism.

In section 1 of the paper, I'll give some background. I'll briefly review the major indeterminacy arguments which are supposed to lend support to predicate naturalism. As we'll see, those indeterminacy arguments come in two forms: (i) subsentential indeterminacy arguments (which are most associated with Putnam) and (ii) sentential indeterminacy arguments (which are most associated with Kripke's Wittgenstein). We'll also consider an attempt by Sider to argue that logical naturalism is required to rebut an extension of Kripke's sentential indeterminacy argument.

For the rest of the paper, I'll present two of my own indeterminacy results, each of which extends Putnam-style subsentential indeterminacy arguments. The first result (section 3) appeals to a weak assumption about the expressive limitations of our language in order to rigorously demonstrate that predicate naturalism is not enough to prevent radical indeterminacy in the meanings of our quantifier, even if it is enough to prevent radical indeterminacy for our predicates. This establishes that predicate naturalism and reference magnetism alone cannot secure determinate reference for our quantifier. To accomplish this, logical naturalism is required. The first result thereby gives us a new path – via purely subsentential indeterminacy – to the same conclusion that Sider argued for using Kripkean considerations.

The second result (section 4) goes one step further – and delivers a more significant and surprising conclusion. Given a slightly stronger assumption, I'll show that, predicate naturalism leaves our logical vocabulary *and our predicates* radically indeterminate. That is, I'll argue that predicate naturalism and reference magnetism alone cannot secure determinate reference *for our predicates*, let alone for our logical vocabulary. To accomplish even this modest goal, logical naturalism is required.

# 1 Extant Indeterminacy Arguments

Broadly speaking, there are two sorts of indeterminacy arguments that have motivated predicate naturalism and reference magnetism.[6] Both target a two-step "top-down" metasemantic theory. According to that theory – which we can call *simple top-down metasemantics* – (i) our dispositions to use certain sentences fully determines the semantic values of those sentences; and (ii) the semantic values for those sentences fully determines the

[6] Much of my understanding of these arguments comes from the work of J. Robert G. Williams. See in particular: J. Robert G. Williams, "Eligibility and Inscrutability," *Philosophical Review*, CXVI, 3 (2007): 361–99; J. Robert G. Williams, "The Price of Inscrutability," *Noûs*, XLII, 4 (2008): 600–41; J. Robert G. Williams, "Lewis on Reference and Eligibility" in Barry Loewer and Jonathan Schaffer, eds., *A Companion to David Lewis* (Malden: Wiley, 2015), pp. 367–382; J. Robert G. Williams, *The Metaphysics of Representation* (Oxford: Oxford University Press, 2020). Note, however, that Williams does *not* accept reference magnetism (although his preferred theory mimics it in certain respects).

semantic values for the primitive vocabulary of our language. The first sort of indeterminacy argument – associated with Kripke's reading of Wittgenstein – wields the threat of *sentential* indeterminacy and targets the first step of simple top-down metasemantics. A second sort – associated with Putnam[7] – wields the threat of *subsentential* indeterminacy and targets the second step of simple top-down metasemantics.

Here is how (a streamlined version of) Kripke-style sentential indeterminacy arguments work. Because our dispositions to use sentences are limited, there are radically different truth-conditions which we can assign to our sentences and which nevertheless best conform to our dispositions. Take, for instance, our dispositions to use the term '$+$' and consider a sentence of the form $\ulcorner n + m = l \urcorner$ where $n$, $m$, and $l$ refer to mind-boggling large numbers. Arguably, we have no dispositions to use such a sentence: it is beyond our discursive abilities to even token it. So, any two interpretations which assign different truth-conditions to such a '$+$' sentence, while agreeing on the truth-conditions for the sentences our dispositions do cover, will fit our dispositions equally well. That is, an interpretation on which $\ulcorner n + m = l \urcorner$ means that $l$ is the result of adding $n$ and $m$ fits our dispositions no better than an interpretation on which the sentence means that $l$ is the result of *quadding* $n$ and $m$, where quadding is the same function as adding for moderately sized numbers, but results in 5 for extremely large numbers.

If the first thesis of simple top-down meta-semantics is true, then, since we've shown that multiple radically different candidate assignments of sentence-level semantic values conform to our dispositions, multiple radically different candidate assignments are equally "correct". We're left with radical indeterminacy as to the meanings of our sentences. That's absurd. Surely $\ulcorner n + m = l \urcorner$ is a claim about addition and not about quaddition. So, it looks like we should reject the first thesis of simple top-down metasemantics: sentence-level semantic values are not entirely determined by our dispositions to use.

Putnam-style subsentential indeterminacy arguments target the second thesis of sim-

[7] In abbreviated re-tellings, Putnam is often credited for subsentential indeterminacy arguments. But the history of such arguments is more complex. That history certainly began before Putnam. Quine's famous inscrutability of reference arguments were an important prelude to the subsentential indeterminacy arguments described here. Indeed, in the late 60s/early 70s, such Quinean arguments led Lewis to doubt whether sentential meanings nail down a unique subsentential assignment. And indeterminacy arguments very much in the style attributed to Putnam – those which permute subsentential meanings while preserving sentential meanings – were explicitly developed in the mid-70s by John Wallace and discussed by Hartry Field and Donald Davidson. (See W.V. Quine, *Word and Object* (Cambridge: MIT Press, 1960); W.V. Quine, *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969); David Lewis, "Languages and Language," in Keith Gunderson, ed., *Minnesota Studies in the Philosophy of Science, Volume VII* (Minneapolis: University of Minnesota Press, 1975), pp 3–35 at pp 20–21; John Wallace, "Only in the Context of a Sentence do Words Have Any Meaning," *Midwest Studies in Philosophy*, II, 1 (1977): 144–164; Hartry H. Field, "Conventionalism and Instrumentalism in Semantics," *Noûs*, IX, 4 (1975): 375–405; Donald Davidson, "The Inscrutibility of Reference," *The Southwestern Journal of Philosophy*, X, 2 (1979): 7–19. Thanks to both referees for helping me fill in some of the history here.)

ple top-down meta-semantics. Begin by assuming that there is a determinate way of assigning semantic values to our sentences. And, assume further that there is a unique "intended" interpretation of our language, which assigns the "correct" semantic values to the primitive vocabulary and generates those sentential semantic values. We then note that there are ways of constructing "twisted" interpretations of that language which assign very different semantic values to the primitive vocabulary but nevertheless assign, for each sentence in the language, the same sentential semantic value assigned by the intended interpretation. On one simple way of constructing the deviant interpretation ("the permutation construction"), we take some permutation $P$ of the objects in the domain, and let the twisted interpretation assign to each primitive vocabulary the image under $P$ of the semantic value assigned by the intended interpretation. On this method, we can construct an interpretation on which, e.g., the extension of "dog" includes some real number, but all sentences have the same truth-conditions. A second way of constructing a deviant interpretation ("the restricted construction") relies on a version of the Löwenheim-Skolem thesis. By that thesis, there is a way of cutting down the uncountable intended domain to a countable set and restricting the intended semantic values of predicates and names, while preserving the truth-conditions of all of the sentences. Interpretations constructed via restriction include ones on which, e.g., the extension of 'real number' is a rather gerrymandered countable subset of the reals.

If the second thesis of simple top-down meta-semantics is true, then since we've shown that *many* radically different candidate interpretations produce the correct sentence-level semantic values, multiple radically different candidate interpretations are equally correct: we're left with radical indeterminacy as to the meanings of our subsentential vocabulary. That's absurd. Surely our term 'dog' does not have, e.g., the number 2 in its extension. So, it looks like we should reject that meta-semantic claim: other factors help fix the correct interpretation of our language besides merely the accuracy with which an interpretation produces sentential semantic values.

Following the lead of Lewis, lots of metaphysicians have seen a role for predicate naturalism in rebutting these indeterminacy arguments.[8] Assume that certain candidate semantic values for our predicates are more-or-less natural. So, for instance, the semantic value that necessarily applies to all and only green things is more natural than the semantic value that necessarily applies to all and only grue things. We can then replace simple

[8]Wolfgang Schwarz has persuasively argued that Lewis doesn't endorse this view, but merely moots it for those that accept certain assumptions of Putnam's. (See Wolfgang Schwarz, "Against Magnetism," *Australasian Journal of Philosophy*, XCII, 1 (2014): 17–36.) Instead, the primary role that naturalness plays in Lewis's considered theory is in fixing our propositional attitudes via considerations of charity. For more on the role naturalness played in Lewis's own theory see also Brian Weatherson, "The Role of Naturalness in Lewis's Theory of Meaning," *Journal for the History of Analytical Philosophy*, I, 10 (2013): 1–19; and J. Robert G. Williams, "Representational Skepticism: The Bubble Puzzle," *Philosophical Perspectives*, XXX 1 (December 2016): 419–42.

top-down metasemantics with the following theory:[9]

**Reference Magnetism** An interpretation is intended to the extent that it best combines *eligibility* and *fit*. An interpretation is *eligible* to the extent that it assigns more natural semantic values to primitive vocabulary, and an interpretation *fits* to the extent that the sentential semantic values it induces conform to our linguistic dispositions.

Predicate naturalism and reference magnetism is thought to rebut Putnam-style subsentential indeterminacy results. An interpretation constructed via permutation on which (for example) the number 2 is in the extension of 'dog' will score as well as the intended interpretation with respect to fit. But it scores poorly with respect to eligibility because this twisted interpretations deploys a semantic value for 'dog' that is less natural than the rival value which necessarily applies to all and only dogs. Similarly with the gerrymandered interpretation of 'real number' constructed via restriction: this twisted semantic value is less natural than the one that necessarily applies to all and only real numbers and so the original intended interpretation is preferred to this twisted interpretation.[10]

Moreover, the view is thought to rebut Kripke-style sentential indeterminacy worries. That's because, even though the deviant "quaddition" interpretation fits our dispositions as well as the intended "addition" interpretation, the former does worse with respect to eligibility. The ability for predicate naturalism (in combination with reference magnetism) to fend off these two sorts of indeterminacy threats has been an influential motivation for predicate naturalism.

So much for the motivation for *predicate* naturalism. What about *logical* naturalism? My aim in this paper is to take the indeterminacy threats that originally motivated predicate naturalism and extend them in ways that predicate naturalism alone cannot help

[9] For some (to my mind, serious) objections to Reference Magnetism, see, among others, Williams, "Eligibility and Inscrutability," *op. cit.*; and John Hawthorne, "Craziness and Metasemantics," *Philosophical Review*, CXVI, 3 (2007): 427–40.

[10] That being said, JRG Williams has pointed out that this use of Reference Magnetism to rebut the restricted construction may be too quick. Here's the problem. Consider a slightly different "twisted" interpretation, which assigns the intended semantic value to predicates like 'real number' but simply restricts the domain of the quantifier to range over the smaller, countable Skolemized domain. The extension of the predicates will include objects outside of the domain, but it's not clear why that should matter. And the reference magnetist cannot complain that the predicates' semantic values on this twisted interpretation are any less natural than the values on the intended interpretation, because they are the same semantic values. (See Williams, *The Metaphysics of Representation,* op. cit., p. 52–53, especially fn. 20 and 21.)

This dialectic deserves to be developed more than I can do here. (Can the reference magnetist complain that the quantifier is being restricted by a gerrymandered property? Are quantifier-restricting properties eligibility-makers of interpretations?) I'll just note that, if Williams' response works, it leads to us to a conclusion similar to that of my first result.

with – ways which (given reference magnetism) we need logical naturalism to rebut.[11] I will be focusing on extensions of Putnam-style subsentential indeterminacy arguments, largely setting aside Kripke-style sentential indeterminacy worries. But before I do so, let me note that Kripke-style sentential indeterminacy concerns *can* be extended to impugn our logical vocabulary. Indeed, Sider does just that. He writes:[12]

> ...what about my use of 'everything' insures that it means universal quantification, rather than something that acts like universal quantification for sentences I have uttered in the past, but behaves bizarrely in new sentences [including ones I have no dispositions towards]?...What rules out rampant semantic indeterminacy for quantifiers is just what rules out such indeterminacy for predicates: reality's structure. Other things being equal, joint-carving interpretations of quantifiers are better interpretations.

I will be arguing for a similar conclusion to Sider's. But we shouldn't rest content with Sider's argument alone. That's because there are various ways an opponent might resist his arguments. In particular, an opponent might seek to fix our sentential semantic values without appealing to logical joints of nature. Here are two ways this might go. First: such an opponent might eschew logical joints but instead appeal to the notion of *idealized dispositions* – roughly, our dispositions were we perfectly rational and had the resources to token any sentence. With this notion, our opponent can perhaps say that fit is a matter of conforming to our idealized dispositions rather than our actual dispositions. Even if Kripke's Wittgenstein is right that our actual linguistic dispositions do not cover every sentence, perhaps idealized linguistic dispositions would cover every sentence. And, if so, perhaps our idealized linguistic dispositions will determine a unique intended interpretation and Sider's indeterminacy worry is rebutted. As a second strategy, an opponent might eschew logical joints but instead appeal to the notion of *natural patterns of actions* – patterns of action that "carve nature at its joints". If so, they should also be comfortable appealing to the notion of *magnetized dispositions* – roughly the dispositions that best combines match with our actual dispositions and natural patterns of action. Then, perhaps our magnetized dispositions will cover every sentence, in which case we can say that fit is a matter of conforming to our magnetized dispositions rather than our actual dispositions.[13]

---

[11]Some philosophers think that predicate naturalism *trivially* entails logical naturalism. According to these philosophers, the semantic values of predicates are of the same type as our logical vocabulary: both predicates and logical vocabulary express properties, the latter merely expresses properties of propositions or second-order properties. So, according to these philosophers, once Lewis has shown us that we need more-or-less natural properties in order to account for more-or-less eligible predicates, we get eligibility of logical vocabulary for free! I have no beef with such philosophers, who already accept my desired conclusion. I'll be speaking to those philosophers who distinguish between eligibility with respect to the type of semantic values for predicates and the type of semantic values for logical vocabulary.

[12]Theodore Sider, "Ontological Realism," *op. cit.*, 406–07.

[13]Interestingly, as Schwarz "Against Magnetism," *op. cit.*, p. 26 points out, Lewis himself appeals to the idea of natural patterns action in his own resolution of the Kripkenstein worry. (Lewis, "New Work for a

Perhaps a Siderian can rebut these objections. But, regardless, it would surely help the logical naturalist's case to have a result that extends Putnam-style subsentential indeterminacy arguments as well as Kripke-style sentential indeterminacy arguments. Recall the difference. Kripke-style results point to rival interpretations that match the intended interpretation for a proper subset of the sentences in our language – those that our dispositions cover. Putnam-style results, however, point to rival interpretations that match the intended interpretation for *all* sentences in the language. So, a Putnam-style result would mean that the threat of radical indeterminacy looms even if idealized or magnetized dispositions fix meanings for every sentence in our language. In the next few sections, I give two such results.

## 2 Core Assumptions

I'll present my Putnam-style indeterminacy results in a first-order language $\mathcal{L}$. The language is fully detailed in appendix A. But a few key details are worth discussing here.

First: The language includes a lambda operator. Where, in classical first-order languages, we have sentences like '$\exists x(Fx \vee Gx)$', in our official language we have sentences like '$\exists \lambda x.(Fx \vee Gx)$'. (Although, where familiarity eases readability, I may abbreviate our official sentences by dropping the lambda operator that appears after an '$\exists$'.) The role of the lambda operator is to bind variables, allowing us to construct complex predicates: for instance, it allows us to convert open formulas to closed predicates. The role of our quantifier is, very roughly, to assert that something satisfies the resulting predicate. In classical languages, these two roles are combined into one: the quantifier both binds variables and says that something satisfies the result.[14]

Working with a lambda operator puts us in good company. Natural language semanticists, when giving a formal semantics for natural languages, posit lambda operators in those languages. One reason they do this is because it allows them to treat the semantic value of sentences like '$\exists \lambda x.(Fx \vee Gx)$' as composed, by functional application, of the semantic values of the quantifier and the lambda expression. In languages without a lambda operator, the semantic value for the quantifier is typically taken to be a domain. But domains can neither apply nor be applied to the semantic value for the variable (which, relative to a variable assignment, is an object in the domain) or the semantic value for an open sentence (which, relative to a variable assignment, is a sentential semantic value).

Theory of Universals," *op. cit.*.) He took adding to be a more natural activity than the activity of quadding, and thus more eligible to be the contents of someone's intentions.

[14] For more on this division of labor, see Jason Turner, "Ontological Nihilism" in Karen Bennett and Dean W. Zimmerman, eds., *Oxford Studies in Metaphysics, Volume 6* (Oxford: Oxford University Press, 2011), pp. 1–50 at section 3.2.

With the addition of the lambda operator, the semantic value of a quantifier can be taken to be a function from the semantic values of predicates like '$\lambda x.(Fx \vee Gx)$' to the semantic value of sentences like '$\exists\lambda x.(Fx \vee Gx)$'. This is important for our purposes: once we stop thinking of the semantic values of quantifiers as domains, and start thinking of them as functions from predicational semantic values to sentential semantic values, this dramatically widens the range of possible semantic values for the quantifier.[15] And with this wider range comes a greater degree of flexibility for finding deviant semantic values for our logical vocabulary.

Second: I'll conform to semantic orthodoxy in taking semantic values to be coarse-grained and intensional. So, for instance, I'll take the semantic values of variables (relative to a variable assignment) to be possibilia. And I'll assume that the semantic values of formulas (i.e. $0$ place predicates) are sets of possible worlds ("propositions") and the semantic values of $1$ place predicates map possibilia to propositions. (More generally, the semantic values of $n$>$0$ place predicates are functions from possibilia to the semantic values of $n-1$ place predicates ("$n$ place intensions").) With the first observation in mind, the semantic value for the quantifier ('$\exists$'), will be a function from $1$ place intensions to propositions, and the semantic values for connectives ('$\vee$', '$\neg$') will be functions from propositions to propositions. Assignments of semantic values to primitive vocabulary will be called *interpretations*. These semantic values combine via functional application to produce semantic values for complex expressions. And I'll denote the semantic value of an expression $\gamma$ (according to an interpretation $I$, relative to a variable assignment $\alpha$) using double brackets: $[\![\gamma]\!]_I^\alpha$.

Third: My language includes names. These names rigidly refer (i.e. names refer to the same object across possible worlds) but are required to refer to an actual object (i.e. I assume there is no name in our language for Socrates' merely possible brother.)

I'll follow the schema described above for Putnam-style arguments. We'll assume that there is an intended interpretation, $I^*$, for our language $\mathcal{L}$. Our challenge, then, is to construct twisted interpretations of $\mathcal{L}$. Importantly, our twisted interpretations must have three key features.

**Feature 1.** *The twisted interpretation and the intended interpretation must assign the same meanings to all sentences in the language.* This ensures that the intended interpretation and the twisted interpretation fit with linguistic usage equally well. And so the predicate naturalist cannot appeal to the fit component of reference magnetism to explain why the twisted interpretation is unintended.

**Feature 2.** *The twisted interpretation must assign to primitive predicates semantic values that are at least as natural as the semantic values assigned by the intended inter-*

[15] This move is inspired by the literature on quantifier variance. See the references contained in Rohan Sud and David Manley, "Quantifier Variance," in Ricki Bliss and J. T. M. Miller, eds., *The Routledge Handbook of Metametaphysics* (New York: Routledge, 2021), pp. 100–17.

*pretation.* This ensures that, from the perspective of a predicate naturalist, the intended interpretation is not more eligible than the twisted interpretation. And so, predicate naturalists cannot appeal to the eligibility component of reference magnetism to explain why the twisted interpretation is unintended.

**Feature 3** *The twisted interpretations will assign strange semantic values to the logical vocabulary of the language.* Say that an interpretation is *logically normal* when it assigns:

- '$\vee$' to the function from two propositions to their disjunction (i.e. their union).
- '$\neg$' to the function from a proposition to its negation (i.e. its complement).
- '$\exists$' to the function from 1 place predicate intensions $i$ to $p$ where $w \in p$ just in case something in $w$ satisfies $i$ at $w$. (i.e. there is some possibilia $x$ in $w$ such that $w \in i(x)$).

The intended interpretation $I^*$, I'll assume, is logically normal. But our twisted interpretations will not interpret the logical vocabulary normally. This ensures that from the perspective of a logical naturalist, the intended interpretation *is* more eligible than the twisted interpretation. And so, logical naturalists *can* appeal to reference magnetism to explain why the twisted interpretation is unintended.

If we can find twisted interpretations with these three features, we will have a case for logical naturalism that is entirely analogous to the original Putnam-style case given on behalf of predicate naturalism. Anyone moved by that argument to accept predicate naturalism should go one step further and accept logical naturalism.

Moreover, to strengthen the result, the twisted interpretations will also give *names* the same semantic values as assigned by the intended interpretation. Thus, those that wish to combine Reference Magnetism with some other meta-semantic theory for names (e.g. a Kripkean causal theory) will still have to contend with our results.

## 3 First Result

For our first result, we'll construct a twisted interpretation $I^{T1}$ of $\mathcal{L}$ which assigns a strange interpretation to '$\exists$', while holding all other semantic values fixed. Our strategy here closely mirrors the strategy of Sider's Kripke-style argument in the following regard. Sider's suggestion was to "bend" the semantic value of '$\exists$' with respect to sentences in $\mathcal{L}$ that we have not uttered and have no dispositions towards – and to do so without affecting how '$\exists$' applies to sentences that our dispositions do cover. We will adopt a similar strategy: we'll bend the semantic value of '$\exists$' with respect to semantic values that are inexpressible in $\mathcal{L}$ – and to do so without affecting how '$\exists$' applies to semantic values that are expressible. Unlike Sider's Kripke-style argument, our twisted interpretation will preserve the semantic value of *every sentence* in $\mathcal{L}$. We'll thereby demonstrate that, even if the predicate naturalists can dodge Sider's sentential indeterminacy threats, they continue to face radical *subsentential* indeterminacy threats – threats which the logical naturalist can avoid. (In comparison, our second result (given in the next section) will follow a very different strategy from this one – and will offer even more significant upshots.)

The first result is based on the following assumption. Say that a proposition is *ineffable* when we cannot express it in our language, under the intended interpretation, with any (open or closed) formula, under any variable assignment. I'll assume:

**First Inexpressibility Assumption (FIA)** There is some ineffable proposition.

FIA is immensely plausible. Many of our terms – e.g. 'knowledge' – cannot be defined in other terms. Let $\mathcal{L}^-$ be the (interpreted) fragment of our language which lacks the word 'knowledge' and other epistemic vocabulary. Presumably the proposition that *Williamson knows that grass is green* is ineffable in $\mathcal{L}^-$. But surely there is a possible language, $\mathcal{L}^+$, that stands to our language as our language stands to $\mathcal{L}^-$. So, surely there is a proposition, expressible in $\mathcal{L}^+$, that is ineffable in our language.

I can think of one strategy for objecting to the above argument for FIA. The strategy is based on anti-nominalism about property talk, according to which our quantifier ranges over properties along with individuals. Note that our notion of ineffability is quite strong. In particular, it goes beyond closed sentences: an ineffable proposition cannot be expressed by any *open* formula under any variable assignment. However, if we reject property nominalism, our variables can be assigned to properties. This would make it extremely easy to express propositions with open formulas. Consider again $\mathcal{L}^-$, which lacks epistemic vocabulary. Let $\mathcal{L}^-$ contain the predicate 'instantiates'. Consider the open formula '$x$ instantiates $y$' and an assignment where '$x$' is assigned to Williamson and '$y$' is assigned to the property in our domain of *knowing that grass is green*. Arguably this formula expresses, under this assignment, the proposition that *Williamson knows that grass is green*. So, *contra* the suggestion above, the proposition is not ineffable in $\mathcal{L}^-$. Analogously, given that our language allows us to construct the same open formula, it's hard to see why any proposition would be ineffable for us.

What should we think of this objection to FIA? Can we still appeal to the assumption in the construction of our deviant interpretation? I am willing to admit that FIA is true *only on the supposition of nominalism about properties*. But I think this observation is a red herring – and not just for those that prefer the desert landscape of nominalism. That's because, intuitively, the threat of radical indeterminacy should not turn on debates over nominalism about properties! It would be bizarre if the fate of reference was held hostage by this relatively recherché ontological debate.[16] So, if reference magnetism together with predicate naturalism is to be an adequate theory of reference, it must be adequate whether

[16] I am assuming that our imagined nominalist can still help themselves to talk of the relative eligibility of the semantic values of predicates, even if they cannot quantify over properties.

or not we can quantify over properties. So, if I can show that the theory is inadequate on the assumption of nominalism, that should make even an anti-nominalist suspicious.

The point is admittedly delicate. (After all, the suggestion that certain properties are reference magnets is also recherché metaphysics.) But I think we can see the irrelevance of anti-nominalism by considering the following thought experiment. Suppose we do in fact quantify over properties and that FIA is false. We can nevertheless imagine a tribe of nominalistic speakers. These speakers eschew quantification over properties for nominalistic paraphrases. The lack predicates like 'instantiates'. And their quantifier (and variable assignments) have a restricted range, a domain that excludes properties. My argument above for FIA goes through for *this* language – there are propositions that are ineffable in their language. And, if logical nominalism is required to secure determinate reference for their language, that's enough to motivate the position. After all, radical indeterminacy of reference in our language is no more absurd than radical indeterminacy in their language!

In sum: we should be happy to assume FIA *for the purposes of testing the need for logical naturalism*. (All that having been said, to ease exposition, I'll help myself to anti-nominalist talk – freely reifying properties – instead of constantly nominalizing this talk.)

Note that one consequence of FIA is:

**Corollary of FIA** Let $p^e$ be an ineffable proposition and let $i^i$ be the following 1 place intension: a function from any possibilia to $p^e$. There is no predicate which expresses $i^i$ (on any variable assignment) – the intension is inexpressible.[17]

With this corollary in hand, we can construct our twisted interpretation $I^{T1}$. The idea is to leave the intended interpretation intact except with respect to the quantifier. On the twisted interpretation, the semantic value of the quantifier acts normally except when combined with our inexpressible intension $i^i$. When combined with that intension, it returns some unexpected, arbitrary proposition (e.g. that monkeys fly). More carefully:

T1-i. Let all primitive vocabulary other than '$\exists$' match the intended interpretation: for all primitive vocabulary $\gamma$ other than '$\exists$', $[\![\gamma]\!]^{\alpha}_{IT1} = [\![\gamma]\!]^{\alpha}_{I*}$.

T1-ii. Let the twisted semantic value of '$\exists$' match the normal, intended, semantic value, except when it combines with $i^i$: For all one place intensions $i$ other than $i^i$, $[\![\exists]\!]^{\alpha}_{IT1}(i) = [\![\exists]\!]^{\alpha}_{I*}(i)$. For $i^i$, '$\exists$' behaves differently: $[\![\exists]\!]^{\alpha}_{IT1}(i^i)$ is some arbitrary proposition $p$, where $p \neq [\![\exists]\!]^{\alpha}_{I*}(i^i)$ (e.g. $p$ can be the proposition that monkeys fly).

[17] *Proof.* Suppose for reductio that there were such a predicate, $\zeta$, which, relative to some assignment $\alpha$, expressed $i^i$. $\zeta$ would have to be a one place predicate and when attached to some variable $\tau$, '$\zeta\tau$' would express the ineffable proposition: $[\![\zeta\tau]\!]^{\alpha}_{I*} = p^e$. Reductio.

It's easy to see that our twisted interpretation has Feature 2 (because the primitive predicates have the same semantic values across both interpretations) and Feature 3 (because '$\exists$' is given a strange interpretation). Moreover, our twisted interpretation has Feature 1: for every sentence in the language, it assigns the same semantic value as the intended interpretation. Here's another way to say the same thing. Say that a formula $\phi$ *is diseased*$_T$ just in case for some assignment $\alpha$ the intended interpretation and the twisted interpretation $I^T$ assign different semantic values to $\phi$. Similarly, say that a predicate $\zeta$ is diseased$_T$ just in case, for some assignment $\alpha$, the intended interpretation and $I^T$ assign different semantic values to $\zeta$. We can prove:

**First Result** No sentence is diseased$_{T1}$: for any sentence $\phi$ and assignment function $\alpha$: $[\![\phi]\!]^{\alpha}_{I^{T1}} = [\![\phi]\!]^{\alpha}_{I^*}$.

A proof is included in appendix B. But the underlying idea is simple enough. Imagine we build formulas and predicates from our primitive vocabulary, producing more and more complex formulas and predicates at each stage by using logical connectives or the lambda operator on less complex expressions from the previous stages. At each stage we can check for disease – check to see if the semantic values of the resulting expressions match across the two interpretations. If we were to ever end up with a diseased formula, the first stage where we would introduce the disease would come when we attach an '$\exists$' to some less-complex predicate, $\zeta$, from previous stages. That's because the two interpretations assign the same semantic values for all logical vocabulary except '$\exists$'. Moreover, $\zeta$ must express $i^i$ on the twisted interpretation, because that's the only place where the two meanings of '$\exists$' diverge. Finally, because this is the *first* stage where we introduce disease, $\zeta$ would be undiseased. But, of course, we know this can't happen. No predicate expresses $i^i$ on the intended interpretation, so no undiseased predicate – including $\zeta$ – expresses $i^i$ on the twisted interpretation.

As with the original Putnam-style indeterminacy arguments, our result puts serious pressure on the second thesis of simple top-down meta-semantics. We've shown that, if the semantic values for our sentences fully determine the semantic values of our subsentential vocabulary, there will be radical subsentential indeterminacy. Of course, there is an interesting difference between my argument and the original Putnam-style indeterminacy arguments. Putnam's twisted interpretations involved making a change to one expression of our language and compensating for that change using another expression. In my argument, I am simply making a change to one expression in the language, '$\exists$', and then relying on the expressive limitations of the language to ensure that this change never affects the truth-conditions of sentences in the language – no compensation necessary. Nevertheless, the upshot is the same: we've found alternative (and obviously incorrect) subsentential assignments that generate the same sentential meanings. And this radical subsentential indeterminacy persists even if we could resolve Kripke (and Sider) style sentential indeterminacy worries – even if our linguistic dispositions did somehow fully determine the semantic values of all sentences in our language.

Continuing the analogy with the original Putnam-style indeterminacy arguments, we might appeal to reference magnetism for help securing determinate reference. But reference magnetism combined with merely predicate naturalism will not help resolve the indeterminacy. That's because the only primitive vocabulary over which the two interpretations differ is *logical* vocabulary, and in particular the term '$\exists$'. In so far as we pin our hopes on reference magnetism to rebut radical subsentential indeterminacy, it looks like we must go beyond *predicate* naturalism and accept *logical* naturalism.

How might a fan of the original Putnam-style indeterminacy argument for predicate naturalism object to my argument for logical naturalism? Such an opponent might be uncomfortable with radical indeterminacy in the semantic value of predicates, without being similarly disturbed by radical indeterminacy in the semantic values of logical vocabulary. Suppose, for instance, we think that the "meaning" of a logical connective is exhausted by its inferential role – the meaning of our logical connectives should be unique *up to inferential role*. In this case, we need not be concerned if the compositional semantic values for our quantifier is indeterminate between several candidates, so long as those candidate preserve all of the same inferential relations between sentences in the language. But, assuming that the inferential relations between sentences is determined by the propositions those sentences express, this is the position we are in with respect to the rival semantic values of '$\exists$'. By the First Result, the twisted interpretation preserves sentence-level semantic values, and so predicts the same inferential relations as the intended interpretation.

I am skeptical of this reply. Our opponent is claiming that the meaning of logical connectives is exhausted by inferential role but does not extend the same courtesy to predicates. That disunified treatment strikes me as objectionably ad hoc. But perhaps my skepticism can be rebutted. Perhaps there are deep differences between logical and non-logical vocabulary that justifies a disunified treatment. Instead of playing out this dialectic further, I'll simply present a second Putnam-style result, which also support logical naturalism, but circumvents these concerns. That's because this second result demonstrates that predicate naturalism alone cannot even secure determinate reference for our predicates, let alone for our logical vocabulary.

# 4 Second Result

Our second result, like the first, consists of constructing a twisted interpretation, $I^{T2}$, that preserves sentence-level semantic values while gerrymandering the values assigned to logical terms. However, in this case, we will *also* modify the meaning of a predicate. A bit more specifically: our twisted interpretation will take a particular "target predicate",

'$F$', and assign it to a strange property that is no less natural than its intended meaning, leaving the other predicates and names alone (this secures Feature 2). And then we make compensating adjustments to the meanings of our logical vocabulary (securing Feature 3) so that sentential meanings are all preserved (securing Feature 1). In this way, our twisted interpretation threatens indeterminacy with respect to our predicates, in addition to our logical vocabulary.

This result will be based on the following assumption, which gives us the deviant semantic value for the target predicate. Let an *alien property* be a property that is not instantiated in the actual world. And recall that an ineffable proposition is one we cannot express in our language, under the intended interpretation, with any (open or closed) formula, under any variable assignment. I will assume:

**Second Inexpressibility Assumption (SIA)** There is an alien perfectly natural property $g^a$ and a corresponding non-empty intension $i^a$ ("the alien intension") such that: (i) for any actual object $x$, $i^a(x)$ is the empty proposition $\emptyset$ (i.e. no actual object instantiates $g^a$ in any world); and (ii) for any possibilia $x$, $i^a(x)$ is either the empty proposition or an ineffable proposition.

At first glance, SIA might look tenuous. But, upon reflection, SIA is quite plausible. Here's one way to see this.

Quarks come in different flavors – up, down, top, bottom, strange and charm. Let's pretend[18] that these flavors are perfectly natural properties and that they can only apply to quarks. Moreover, let's suppose that flavors are contingent properties (physicists talk of quarks changing flavors) and that being a quark is an essential property (something that is a quark couldn't have been, say, an electron). Take the perspective of a language community in a world $w$ that lacks quarks and speaks a language $\mathcal{L}^-$ which lacks the relevant physical vocabulary (e.g. lacks the terms 'quark', 'up', etc.). From their perspective, consider the perfectly natural property of being flavored up and the corresponding intention $i^u$. Because $w$ contains no quarks (i.e. being a quark is "alien" relative to $w$) and being a quark is an essential property, no object in $w$ is a quark in any world. And because only quarks can be up flavored, no object in $w$ is up flavored in any world. That is, for any object $x$ in $w$, $i^u(x)$ is the empty proposition. Now consider some possibilia $x$ outside of $w$. If $x$ isn't a quark at any world, $i^u(x)$ is the empty proposition (because only quarks can be up flavored). If $x$ is a quark at some world, $i^u(x)$ is very plausibly ineffable in $\mathcal{L}^-$. That's because the language lacks the requisite physical vocabulary to distinguish, say, a world where $x$ exists alone and is flavored up from a world where $x$ exists alone but is flavored down.

[18] Emphasis on 'pretend': I've been warned that philosophers of physics can find it irritating when metaphysicians claim quark properties are fundamental. I hope they will forgive this metaphysician for engaging in some illustrative pretense!

In other words, from the perspective of this merely possible linguistic community, SIA is true. But surely there is a language and world, with a collection of perfectly natural alien properties (e.g. *shavors* of *shuarks*) which stands to our language and world as our language and world stands to $\mathcal{L}^-$ and $w$. Consider one such perfectly natural alien property, e.g. *being shup*, and the corresponding intention $i^s$. Because the actual world contains no shuarks and because being a shuark is (like being a quark) an essential property, no actual object is a shuark in any world. And, because only shuarks can be shup shavored, no actual object is shup at any world – that is, for any actual object $x$, $i^s(x)$ is the empty proposition. Now consider some possibilia $x$ outside of the actual world. If $x$ isn't a shuark in any world, $i^s(x)$ is the empty proposition (because only shuarks can be shup shavored). If $x$ is a shuark in some world, $i^s(x)$ is ineffable in our language because we lack the requisite "shysical" vocabulary.

As with my defense of FIA, my defense of SIA involves some controversial metaphysical assumptions (e.g. that quarks are essentially quarks). And those that reject these metaphysical assumptions might see a path to resist my argument. But, once again, I think this is a red herring: it would be objectionably bizarre if our theory of reference turned on whether or not, say, quarks are essentially quarks! So, we should be happy to assume SIA, at least for the purposes of testing the need for logical naturalism.

In addition to SIA, I will also assume that there is a primitive predicate that has the empty intension (i.e. it returns the empty set no matter what is fed into it) under the intended interpretation (e.g. 'fictional', 'imaginary', 'non-self-identical', or (if fictional species necessarily fail to exist) 'unicorn'.) (I make no assumption about how natural this intension is.) This predicate will serve as the *target predicate* that our twisted interpretation will seek to bend.[19] I'll let '$F$' in our formal language be this target predicate.

We can now begin to define a twisted interpretation, starting with clauses for non-logical vocabulary (names and primitive predicates). Here, the twisted interpretation will match the intended interpretation except in one crucial respect: the target predicate will be assigned to the alien intension. All other predicates and all names have the same interpretation as the intended interpretation. Because $i^a$ is perfectly natural, our twisted interpretation has Feature 2: without logical naturalism, the twisted interpretation is no less eligible than our intended interpretation. (Indeed, if the empty intension is less than perfectly natural, the semantic value of '$F$' will be *more* eligible on the twisted interpretation than on the intended interpretation!) Our challenge is to continue constructing the twisted interpretation so that sentences have the same semantic value on the twisted interpretation as they do on the intended interpretation (Feature 1) while "bending" the meaning of logical vocabulary (Feature 3). Is this possible?

From what we've said about the non-logical vocabulary, we can already see that the

[19] I don't mean to suggest that this is the only predicate that we can bend. As far as I can tell, other predicates can be bent as well, although they will require different, more complex constructions.

twisted interpretation will induce deviations in the meanings of at least *some* expressions. That is, the clauses for non-logical vocabulary mean that the following expressions (and anything equivalent[20] to them) are diseased$_{T2}$:

- *The target predicate '$F$'*: $[\![F]\!]^{\alpha}_{I*}$ is the empty intension, while $[\![F]\!]^{\alpha}_{IT2}$ is the alien intension. Similarly, repeated lambda-abstractions on '$F$' are also diseased: $\ulcorner\lambda\tau_1...\lambda\tau_n.F\urcorner$.

- *Open formulas of the form* $\ulcorner F\upsilon\urcorner$: $[\![Fx]\!]^{\alpha}_{I*}$ is always the empty proposition, while $[\![Fx]\!]^{\alpha}_{IT2}$ is sometimes (i.e on some assignment functions) an ineffable proposition. Similarly, repeated lambda-abstractions on such formulas are also diseased: $\ulcorner\lambda\tau_1...\lambda\tau_n.F\upsilon\urcorner$.

Let's say that these are the *initially infected* expressions. Importantly, we can live with this initial infection. That's because none of the initially infected expressions are *sentences* (i.e. closed formula) and so they don't undermine Feature 1. So if we can bend the semantic values for the logical vocabulary so that *only* these initially infected expressions are diseased, we will have secured all three features desired for our twisted interpretation. Can this be done?

To start, we must make sure that the disease does not spread from the initially infected expressions to new predicates and formulas built from them. There are three ways we can build new expressions – three ways our initial infection might spread.

1. We can build by *abstraction* by taking a diseased expression and prefixing a lambda operator and variable to it. For instance, we can take the diseased expression '$Fx$' and abstract to get '$\lambda x.Fx$'.

2. We can build by *application* by applying an initially infected predicate to a variable or a name. For instance, we can take the diseased predicate '$F$' and apply it to a name '$a$' or variable '$x$' to get '$Fa$' or '$Fx$'.

3. We can build by *logical construction* by applying a logical connective to an initially infected expression. For instance, we can apply the connective '$\neg$' to '$Fx$' to get '$\neg Fx$'. Or we can apply '$\exists$' to '$\lambda x.Fx$' to get '$\exists xFx$'.

If we are to contain the disease, we need to make sure that by using these three ways of building, we either stay within the set of initially infected expressions or else end up with an expression that is no longer diseased.

*Abstraction.* This is easy to see for expressions that result from abstraction on the initially infected expressions. Because the set of initially infected expressions is closed under abstraction, we know that abstraction never spreads disease.

[20] The notion of equivalence I have in mind is $\beta\eta$-equivalence, defined in appendix A.

*Application.* Applying variables to an initially infected predicate will result in one of the initially infected expressions. For instance, applying '$x$' to '$F$' results in '$Fx$', which is one of the initially infected expressions. And applying '$y$' to '$\lambda x.Fx$' results in (an expression equivalent to) '$Fy$', and so is among the initially infected expressions.

What about applying names like $a$ to the initially infected expressions? Sometimes, applying names will be trivial in the sense that it is equivalent to an expression that doesn't contain a name. For instance applying '$a$' to '$\lambda x.Fy$' results in (an expression equivalent to) $Fy$. In such cases, applying names will (like applying variables) simply result in one of the initially infected expressions. But sometimes applying names to an initially infected expression will be non-trivial. And in such cases, the result is an expression that is outside the set of the initially infected expressions. For instance, applying '$a$' to '$F$' results in '$Fa$' which is not initially infected. Fortunately, in such cases, the resulting expression is not diseased. Such cases involve the name combining with the target predicate '$F$'. But the target predicate '$F$' treats named objects the same across either interpretation. That's because names pick out actual objects (we can't name Socrates' merely possible brother) and $[\![F]\!]^{\alpha}_{I^*}$ (the empty intension) and $[\![F]\!]^{\alpha}_{IT2}$ (the alien intension) both map any actual object to the empty proposition. Upshot: applying a name in a way that takes us outside the set of initially infected expressions results in a expression that is no longer diseased. In slogan form: names cure the initial infection!

*Logical construction.* Once we combine our initially infected expressions with logical connectives, we produce expressions like '$\exists x Fx$' which are not among the initially infected ones. So we need to "bend" our logical vocabulary in such a way that, when it combines with any of the diseased expressions, the resulting expression is cured: it returns an undiseased expression.

The fully specified clauses for the logical vocabulary are given in appendix C. But here's the underlying idea. Consider the initially infected expressions which can combine with our logical connectives – 1 place predicates like '$F$' and '$\lambda x.Fy$' and formulas like '$Fx$'. We simply interpret our logical vocabulary so that they treat these expressions *as if* they had their intended meanings. Take the expression '$F$'. Because '$F$' expresses the alien intension on the twisted interpretation (instead of the intended empty intension) we simply bend '$\exists$' so that it treats the alien intension as if it were the empty intension. In other words, $[\![\exists]\!]^{\alpha}_{IT2}$ maps the alien intension to the empty proposition. Similarly for the open formula '$Fx$' and the other connectives: because '$Fx$' sometimes expresses an ineffable proposition (instead of the intended empty proposition) we bend '$\vee$' and '$\neg$' to treat such ineffable propositions as if they were the empty propositions. Finally take the strange intension sometimes expressed by '$\lambda y.Fx$', which maps possibilia to an ineffable proposition. We simply bend '$\exists$' so that it also treats this strange intension as if it were the intended empty intension: $[\![\exists]\!]^{\alpha}_{IT2}$ maps this strange intension to the empty proposition.

By bending our logical vocabulary in this way, we ensure that the disease doesn't spread from the initially infected expressions via logical constructions. We also secure

Feature 3 for our twisted interpretation: our logical vocabulary is not normal.

We've just seen that the disease won't spread from the initially infected expressions to expressions built from the initially infected expressions. There remains the possibility that some other sentence – one not built from the initially infected expressions – is diseased. In fact this doesn't occur. As I prove in appendix C, *only* these initially infected expressions are diseased. That's to say: the diseased predicates and formulas of the language are completely contained by the initially infected expressions. And, as a result, we can see that no sentence is diseased:

**Second Result** For any sentence $\phi$: $[\![\phi]\!]_{I^{T2}} = [\![\phi]\!]_{I^*}$

So, our twisted interpretation has Feature 1 and thus all three of the desired features of a twisted interpretation. Note, importantly, that unlike our previous result, this result impugns the determinacy of our predicates – in particular, our target predicate '$F$' – in addition to our logical vocabulary. The upshot is that, assuming reference magnetism, predicate naturalism alone isn't enough to secure determinate reference even for our predicates, let alone for our logical vocabulary. For even this modest goal, we need logical naturalism.

***

With these two results in hand, I conclude that, in so far as we were moved by Putnam's original indeterminacy argument to accept predicate naturalism, we should be moved to accept logical naturalism. Without logical naturalism, reference magnetism fails to secure determinate reference for our logical vocabulary (First Result) and – much more surprisingly – fails to secure determinate reference for our predicates (Second Result).

# A Syntax and Semantics

Let $\mathcal{L}$ be a first-order language with a lambda operator. The primitive vocabulary includes $n$ place ($n \geq 0$) primitive predicates ('$F$', '$G$',...); sentential connectives '$\vee$' and '$\neg$'; a predicate operator '$\exists$'; a lambda operator '$\lambda$'; variables ('$x$','$y$',...); names ('$a$','$b$',...); and parentheses. We can define the notions of a *singular term* and an ($n$ place) *predicate* in $\mathcal{L}$ as follows:

- Singular Terms: All variables and names are *singular terms* in $\mathcal{L}$; nothing else is a singular term in $\mathcal{L}$.
- Predicates

- All primitive $n$ place predicates are $n$ place predicates in $\mathcal{L}$.
- If $\phi$ is an $n$ place predicate and $\tau$ is a variable, then $(\lambda\tau.\phi)$ is an $n$+1 place predicate.
- If $\phi$ is an $n$>0 place predicate and $\sigma$ is a singular term, then $(\phi\sigma)$ is an $n$-1 place predicate.
- If $\phi$ is a 1 place predicate, then $(\exists\phi)$ is a 0 place predicate.
- If $\phi$ and $\psi$ are 0 place predicates, then $(\neg\phi)$ and $(\phi\vee\psi)$ are 0 place predicates.
- Nothing else is a predicate in $\mathcal{L}$.

A *formula* is a 0 place predicate. The notion of a *free* and *bound* variable is defined in the normal way, with the $\lambda$-operator binding variables. A formula with no unbound variables is a *sentence*; otherwise it's an *open formula*.

Let's turn to the models for $\mathcal{L}$. I'll follow orthodoxy in quantifying over possibilia when giving a semantics for $\mathcal{L}$ in our meta-language. Let a *variable assignment*, $\alpha$, be a function from variables to possibilia. Call a set of possible worlds a *proposition*. Let a 0 *place predicate intension* be a proposition and a $n$>0 *place predicate intension* be a function from objects to an $n$-1 place predicate intension.

An *interpretation function* $I$ assigns, relative to a variable assignment $\alpha$, semantic values to the primitive vocabulary (written $[\![.]\!]_I^\alpha$) as follows:

- **Names**: For names $\pi$, $[\![\pi]\!]_I^\alpha$ is some actual object.
- **Variables**: For variables $\tau$, $[\![\tau]\!]_I^\alpha = \alpha(\tau)$.
- **Primitive Predicates**: For $n$ place primitive predicate $\phi$, $[\![\phi]\!]_I^\alpha$ is some $n$ place predicate intension.
- **Quantifier Symbol**: $[\![\exists]\!]_I^\alpha$ is a function from 1 place predicate intensions to propositions.
- **Negation Symbol**: $[\![\neg]\!]_I^\alpha$ is a function from propositions to propositions.
- **Disjunction Symbol**: $[\![\vee]\!]_I^\alpha$ is a two place function from two propositions to propositions.
- For all primitive vocabulary other than the variables, the semantic value does not vary depending on the variable assignment.

These assignment to primitive vocabulary induce assignments to complex expressions as follows:

- **Atomic Expressions**: For predicate $\phi$ and singular term $\sigma$: $[\![\phi\sigma]\!]^{\alpha}_{I}$ = $[\![\phi]\!]^{\alpha}_{I}$ ($[\![\sigma]\!]^{\alpha}_{I}$).
- **Lambda-Terms**: For variable $\tau$ and $n$ place predicate $\phi$: $[\![\lambda\tau.\phi]\!]^{\alpha}_{I}$ is an $n$+1 place predicate intension that maps possibilia $x$ to $[\![\phi]\!]^{\alpha[\tau=x]}_{I}$, where $\alpha[\tau = x]$ is the assignment function that assigns $\tau$ to $x$ (and is the same as $\alpha$ for all other variables).
- **Complex Formulas**:
  - $[\![\exists\zeta]\!]^{\alpha}_{I} = [\![\exists]\!]^{\alpha}_{I}([\![\zeta]\!]^{\alpha}_{I})$
  - $[\![\phi \vee \psi]\!]^{\alpha}_{I} = [\![\vee]\!]^{\alpha}_{I}([\![\phi]\!]^{\alpha}_{I}, [\![\psi]\!]^{\alpha}_{I})$
  - $[\![\neg\phi]\!]^{\alpha}_{I} = [\![\neg]\!]^{\alpha}_{I}([\![\phi]\!]^{\alpha}_{I})$

Note that the standard equivalence rules are sound in our model theory.

**Definition** (Equivalence)**.** *For predicate $\phi$, singular term $\sigma$, and variable $\tau$, let $\phi[\sigma/\tau]$ be the result of substituting $\sigma$ for every free occurrence of $\tau$ in $\phi$ (and, where necessary, changing bound variables in $\phi$ to prevent $\sigma$ from being bound upon replacement).*[21]

- *Two predicates are immediately $\alpha$-equivalent when either one results from the other by replacing a predicate of the form $\lambda\tau_1.P$ with $\lambda\tau_2.P[\tau_2/\tau_1]$, where $\tau_2$ is not free in $P$.*
- *Two predicates are immediately $\eta$-equivalent when either one results from the other by replacing $\lambda\tau.P\tau$ with $P$, where $\tau$ is not free in $P$.*
- *Two predicates are immediately $\beta$-equivalent when either one results from the other by replacing a predicate $(\lambda\tau.P)\sigma$ with $P[\sigma/\tau]$.*
- *$M$ is $\beta\eta$-equivalent to $N$ ($M \sim_{\beta\eta} N$) when there are a series of predicates $M, P_1, ..., P_n, N$ such that any two adjacent predicates are $\alpha$- $\eta$- or $\beta$-equivalent.*

**Lemma** (Soundness of $\beta\eta$-Equivalence)**.** *If $M \sim_{\beta\eta} N$, then for any interpretation $I$, and assignment function $\alpha$, $[\![M]\!]^{\alpha}_{I} = [\![N]\!]^{\alpha}_{I}$*

*Proof.* It suffices to show that the relevant replacements preserve semantic values. That is, we need to show that, for any interpretation $I$, and assignment function $\alpha$:

1. $[\![\lambda\tau_1.P]\!]^{\alpha}_{I} = [\![\lambda\tau_2.P[\tau_2/\tau_1]]\!]^{\alpha}_{I}$ where $\tau_2$ is not free in $P$.
2. $[\![P]\!]^{\alpha}_{I} = [\![\lambda\tau.P\tau]\!]^{\alpha}_{I}$ , where $\tau$ is not free in $P$.

[21] A precise definition is given in J. Roger Hindley and Jonathan P. Seldin, *Lambda-Calculus and Combinators, an Introduction* (Cambridge: Cambridge University Press, 2008), at p. 7.

3. $[\![(\lambda\tau.P)\sigma]\!]_I^\alpha = [\![P[\sigma/\tau]]\!]_I^\alpha$.

We can see this as follows:

1. Assume $\tau_2$ is not free in $P$. Then $[\![\lambda\tau_1.P]\!]_I^\alpha = x \mapsto [\![P]\!]_I^{\alpha[\tau_1=x]} = x \mapsto [\![P[\tau_2/\tau_1]]\!]_I^{\alpha[\tau_2=x]} = [\![\lambda\tau_2.P[\tau_2/\tau_1]]\!]_I^\alpha$

2. Assume $\tau$ is not free in $P$. Then $[\![\lambda\tau.P\tau]\!]_I^\alpha = x \mapsto [\![P\tau]\!]_I^{\alpha[\tau=x]} = x \mapsto [\![P]\!]_I^{\alpha[\tau=x]}([\![\tau]\!]_I^{\alpha[\tau=x]}) = x \mapsto [\![P]\!]_I^\alpha(x) = [\![P]\!]_I^\alpha$.

3. Assume $\sigma$ is a variable.

$$\begin{aligned}
[\![(\lambda\tau.P)\sigma]\!]_I^\alpha &= [\![(\lambda\tau.P)]\!]_I^\alpha([\![\sigma]\!]_I^\alpha) \\
&= [\![(\lambda\tau.P)]\!]_I^\alpha(\alpha(\sigma)) \\
&= x \mapsto [\![P]\!]_I^{\alpha[\tau=x]}(\alpha(\sigma)) \\
&= [\![P]\!]_I^{\alpha[\tau=\alpha(\sigma)]} \\
&= [\![P[\sigma/\tau]]\!]_I^\alpha
\end{aligned}$$

Assume instead that $\sigma$ is a name.

$$\begin{aligned}
[\![(\lambda\tau.P)\sigma]\!]_I^\alpha &= [\![(\lambda\tau.P)]\!]_I^\alpha([\![\sigma]\!]_I^\alpha) \\
&= [\![(\lambda\tau.P)]\!]_I^\alpha([\![\sigma]\!]_I^\alpha) \\
&= x \mapsto [\![P]\!]_I^{\alpha[\tau=x]}([\![\sigma]\!]_I^\alpha) \\
&= [\![P]\!]_I^{\alpha[\tau=[\![\sigma]\!]_I^\alpha]} \\
&= [\![P[\sigma/\tau]]\!]_I^\alpha
\end{aligned}$$

$\square$

# B First Result

Recall that a predicate (or formula) $\phi$ is diseased$_T$ just in case, for some assignment $\alpha$, the intended interpretation and $I^T$ assign different semantic values to $\phi$. And recall the interpretation $I^{T1}$ (p. 11). We will prove:

**First Result** No sentence is diseased$_{T1}$: for any sentence $\phi$ and assignment function $\alpha$: $[\![\phi]\!]_{I^{T1}}^\alpha = [\![\phi]\!]_{I^*}^\alpha$.

Say that the *complexity* of a predicate (including formulas) is the number of logical constants ($\exists$, $\vee$, $\neg$) or lambda operators ($\lambda$) it contains. Our proof will be by induction on complexity of predicates.

**Base Case.** *Predicates of complexity* $0$ *are not diseased.*

*Proof.* Let $\zeta$ be a predicate of complexity $0$. So $\zeta$ is a $n \geq 0$ place primitive predicate followed by some (possibly 0) singular terms. It follows immediately that $\zeta$ is not diseased. That's because $I^*$ and $I^{T1}$ both assign (relative to a variable assignment) the same semantic values to primitive predicates and singular terms. □

**Induction Step.** *(Logical Operators Don't Cause Disease) Suppose predicates of complexity less than* $n$ *are not diseased. Predicates of complexity* $n$ *are also not diseased.*

*Proof.* Let $\zeta$ be a predicate of complexity $n > 0$. Then it has one of the following forms, where $\phi$ and $\psi$ are formulas, $\chi_m$ is an $m \geq 0$ place predicate, $\tau$ is a variable, and $\sigma_1 ... \sigma_o$ are singular terms:

1. $\neg\phi$
2. $\phi \vee \psi$
3. $(\lambda\tau.\chi_m)\sigma_1...\sigma_o$
4. $\exists\chi_1$

By the inductive hypothesis, $\phi$, $\psi$, and $\chi_m$ are not diseased. Given this, it's easy to see that, for each predicate of form (1)-(4), it is also not diseased:

1. Immediate (because $[\![\neg]\!]^{\alpha}_{I^{T1}} = [\![\neg]\!]^{\alpha}_{I^*}$).
2. Immediate (because $[\![\vee]\!]^{\alpha}_{I^{T1}} = [\![\vee]\!]^{\alpha}_{I^*}$).
3. Because singular terms are not diseased, it suffices to show that $\lambda\tau.\chi_m$ is not diseased. Because $\chi_m$ is not diseased, $[\![\chi_m]\!]^{\alpha}_{I^{T1}} = [\![\chi_m]\!]^{\alpha}_{I^*}$ for *any* assignment function $\alpha$. That includes any variant of an assignment function. So, for any possibilia $x$, $[\![\chi_m]\!]^{\alpha[\tau=x]}_{I^{T1}} = [\![\chi_m]\!]^{\alpha[\tau=x]}_{I^*}$. Thus, $[\![\lambda\tau.\chi_m]\!]^{\alpha}_{I^{T1}} = [\![\lambda\tau.\chi_m]\!]^{\alpha}_{I^*}$, for any assignment $\alpha$.
4. Recall that, by **Corollary of FIA**, no predicate expresses $i^i$ on the intended interpretation. So $\chi_1$ doesn't express $i^i$ on the intended interpretation. Because $\chi_1$ isn't diseased, it doesn't express $i^i$ on $I^{T1}$ either. Because $[\![\exists]\!]^{\alpha}_{I^{T1}}$ differs from $[\![\exists]\!]^{\alpha}_{I^*}$ only with respect to $i^i$, $[\![\exists\chi_1]\!]^{\alpha}_{I^{T1}} = [\![\exists\chi_1]\!]^{\alpha}_{I^{I^*}}$. □

# C Second Result

Recall (p. 14) the alien intension $i^a$ posited in SIA. Using this intension, let's explicitly specify our twisted interpretation $I^{T2}$.

T2-i. **Names** For any name $\pi$: $[\![\pi]\!]^{\alpha}_{I^{T2}} = [\![\pi]\!]^{\alpha}_{I^*}$.

T2-ii. **Primitive Predicates** For the target 1 place predicate '$F$', let the twisted interpretation assign the alien intension: $[\![F]\!]_{I^{T2}} = i^a$. For all other primitive predicates $\zeta$, let the twisted interpretation match the intended interpretation: $[\![\zeta]\!]_{I^{T2}} = [\![\zeta]\!]_{I^*}$.

Say that a 1 place intension $i^{ti}$ is *trivially ineffable* when there is some ineffable proposition $p$ such that for any $x$, $i^{ti}(x) = p$ (i.e. there is some ineffable proposition which the intension spits out regardless of what is feed into it). Let $f$ map the alien intension $i^a$ and any trivially ineffable intensions $i^{ti}$ to the empty intension $i^e$ (i.e. the intension that returns the empty proposition $\emptyset$ no matter what is fed into it) and all other intensions to themselves. And let $f$ map all ineffable propositions to $\emptyset$ and all other propositions to themselves. Then we can say:

T2-iii. **Quantifier Symbol:** For any 1 place intension $i$, $[\![\exists]\!]_{I^{T2}}(i) = [\![\exists]\!]_{I^*}(f(i))$.

T2-iv. **Negation Symbol:** For any proposition $p$: $[\![\neg]\!]_{I^{T2}}(p) = [\![\neg]\!]_{I^*}(f(p))$

T2-v. **Disjunction Symbol:** For any propositions $p, q$: $[\![\vee]\!]_{I^{T2}}(p, q) = [\![\vee]\!]_{I^*}(f(p), f(q))$

We need to prove:

**Second Result** No sentence is diseased$_{T2}$: for any sentence $\phi$ and assignment function $\alpha$: $[\![\phi]\!]^{\alpha}_{I^{T2}} = [\![\phi]\!]^{\alpha}_{I^*}$.

We'll show our Second Result by proving something stronger:

**Theorem.** *For all predicates (or formulas) $\zeta$, at least one of the following holds:*

*1*. $\zeta$ is not diseased.*

*2*. $\zeta \sim_{\beta\eta} \lambda\tau_1.\lambda\tau_2....\lambda\tau_n.F\upsilon$, for variables $\tau_1, ..., \tau_n$ ($n \geq 0$) and variable $\upsilon$.*

*3*. $\zeta \sim_{\beta\eta} \lambda\tau_1.\lambda\tau_2....\lambda\tau_n.F$, for variables $\tau_1, ..., \tau_n$ ($n \geq 0$).*

*(Note that cases 2* and 3* include the cases where $n = 0$: $\zeta \sim_{\beta\eta} F\upsilon$ and $\zeta \sim_{\beta\eta} F$.)*

Once again, our proof will be by induction on complexity of formulas.

**Base Case.** *Let $\zeta$ be a predicate of complexity $0$. One of 1*-3* holds for $\zeta$.*

*Proof.* $\zeta$ is a $m \geq 0$ place primitive predicate $\chi$ followed by $o$ singular terms ($m \geq o \geq 0$). If $\chi$ is not $F$, 1* immediately holds for $\zeta$ (because singular terms and all other primitive predicates have the same interpretation). If $\chi$ is $F$, then $\zeta$ is either of the form:

(i) $F$;

(ii) $Fv$ for some variable $v$; or

(iii) $F\pi$ for some name $\pi$.

If (i), 3* immediately holds for $\zeta$. If (ii), 2* immediately holds for $\zeta$. If (iii), 1* holds for $\zeta$ because $[\![F\pi]\!]^{\alpha}_{I^*} = \emptyset$ ($F$ picks out the empty intension on $I^*$) and $[\![F\pi]\!]^{\alpha}_{IT2} = \emptyset$ (names pick out actual objects on the intended interpretation and no actual object instantiates the alien property in any possible world). $\square$

**Induction Step.** *(Logical Operators Cure Disease) Suppose one of 1*-3* holds for predicates of complexity less than $n$. One of 1*-3* holds for predicates of complexity $n$.*

*Proof.* Let $\zeta$ be a predicate of complexity $n > 0$. Then it has one of the following forms, where $\phi$ and $\psi$ are formulas, $\chi_m$ is an $m \geq 0$ place predicate, $\tau$ is a variable, and $\sigma_1 ... \sigma_o$ are singular terms:

(i) $\neg\phi$

(ii) $\phi \vee \psi$

(iii) $\exists\chi_1$; or

(iv) $(\lambda\tau_*.\chi)\sigma_1 ... \sigma_o$

We consider each case in turn:

(i) Suppose $\zeta$ is $\neg\phi$. By the induction hypothesis, either 1* or 2* holds of $\phi$. (3* cannot hold because $\phi$ is a formula and not equivalent to a $m > 0$ place predicate.) We'll show that, either way, $f([\![\phi]\!]^{\alpha}_{IT2}) = [\![\phi]\!]^{\alpha}_{I^*}$, and use this to show that $\neg\phi$ is not diseased.

- Suppose 1* holds of $\phi$ and $\phi$ is not diseased. So $[\![\phi]\!]^{\alpha}_{IT2} = [\![\phi]\!]^{\alpha}_{I^*}$. So $[\![\phi]\!]^{\alpha}_{IT2}$ is not an ineffable proposition (because $I^*$ cannot pick out an ineffable proposition). So, $f([\![\phi]\!]^{\alpha}_{IT2}) = [\![\phi]\!]^{\alpha}_{IT2} = [\![\phi]\!]^{\alpha}_{I^*}$.
- Suppose 2* holds of $\phi$. Then $\phi \sim_{\beta\eta} Fv$. Depending on the assignment, $[\![Fv]\!]^{\alpha}_{IT2}$ is $\emptyset$ or an ineffable proposition. Either way, $f([\![Fv]\!]^{\alpha}_{IT2}) = \emptyset$. And, of course, $[\![Fv]\!]^{\alpha}_{I^*} = \emptyset$. So, $f([\![\phi]\!]^{\alpha}_{IT2}) = f([\![Fv]\!]^{\alpha}_{IT2}) = \emptyset = [\![Fv]\!]^{\alpha}_{I^*} = [\![\phi]\!]^{\alpha}_{I^*}$.

Given that $f(\llbracket\phi\rrbracket^{\alpha}_{I^{T2}}) = \llbracket\phi\rrbracket^{\alpha}_{I^*}$, we can see $\neg\phi$ is not diseased:

$$\begin{aligned}
\llbracket\neg\phi\rrbracket^{\alpha}_{I^{T2}} &= \llbracket\neg\rrbracket^{\alpha}_{I^{T2}}(\llbracket\phi\rrbracket^{\alpha}_{I^{T2}}) \\
&= \llbracket\neg\rrbracket^{\alpha}_{I^*}(f(\llbracket\phi\rrbracket^{\alpha}_{I^{T2}})) && \text{by the definition of } I^{T2} \\
&= \llbracket\neg\rrbracket^{\alpha}_{I^*}(\llbracket\phi\rrbracket^{\alpha}_{I^*}) \\
&= \llbracket\neg\phi\rrbracket^{\alpha}_{I^*}
\end{aligned}$$

(ii) Suppose $\zeta$ is $\phi \vee \psi$. By the same reasoning as the previous case, $f(\llbracket\phi\rrbracket^{\alpha}_{I^{T2}}) = \llbracket\phi\rrbracket^{\alpha}_{I^*}$ and $f(\llbracket\psi\rrbracket^{\alpha}_{I^{T2}}) = \llbracket\psi\rrbracket^{\alpha}_{I^*}$. Thus $\phi \vee \psi$ is not diseased:

$$\begin{aligned}
\llbracket\phi \vee \psi\rrbracket^{\alpha}_{I^{T2}} &= \llbracket\vee\rrbracket^{\alpha}_{I^{T2}}(\llbracket\phi\rrbracket^{\alpha}_{I^{T2}}, \llbracket\psi\rrbracket^{\alpha}_{I^{T2}}) \\
&= \llbracket\vee\rrbracket^{\alpha}_{I^*}(f(\llbracket\phi\rrbracket^{\alpha}_{I^{T2}}), f(\llbracket\psi\rrbracket^{\alpha}_{I^{T2}})) && \text{by the definition of } I^{T2} \\
&= \llbracket\vee\rrbracket^{\alpha}_{I^*}(\llbracket\phi\rrbracket^{\alpha}_{I^*}, \llbracket\phi\rrbracket^{\alpha}_{I^*}) \\
&= \llbracket\phi \vee \psi\rrbracket^{\alpha}_{I^*}
\end{aligned}$$

(iii) Suppose $\zeta$ is $\exists\chi_1$. By the induction hypothesis, one of 1*-3* holds of $\chi_1$. In any case, $f(\llbracket\chi_1\rrbracket^{\alpha}_{I^{T2}}) = \llbracket\chi_1\rrbracket^{\alpha}_{I^*}$:

- Suppose 1* holds and $\chi_1$ is not diseased. We know that $\llbracket\chi_1\rrbracket^{\alpha}_{I^*}$ is neither the alien intension $i^a$ nor a trivial ineffable intension. Because $\llbracket\chi_1\rrbracket^{\alpha}_{I^{T2}} = \llbracket\chi_1\rrbracket^{\alpha}_{I^*}$, we can conclude that $\llbracket\chi_1\rrbracket^{\alpha}_{I^{T2}}$ also is neither the alien intension $i^a$ nor a trivial ineffable intension. $f$ maps all other intensions to themselves. So, $f(\llbracket\chi_1\rrbracket^{\alpha}_{I^{T2}}) = \llbracket\chi_1\rrbracket^{\alpha}_{I^*}$.
- Suppose 2* holds. Because $\chi_1$ is a 1 place predicate, $\chi_1 \sim_{\beta\eta} \lambda\tau.F\upsilon$ and so $\llbracket\chi_1\rrbracket^{\alpha}_{I^{T2}} = \llbracket\lambda\tau.F\upsilon\rrbracket^{\alpha}_{I^{T2}}$. Depending on the assignment function and whether $\upsilon = \tau$, $\llbracket\lambda\tau.F\upsilon\rrbracket^{\alpha}_{I^{T2}}$ is either the alien intension (when $\upsilon = \tau$), the empty intension (when $\upsilon \neq \tau$ and $\alpha(\upsilon)$ doesn't instantiate the alien property in any world) or is a trivially ineffable intension (when $\upsilon \neq \tau$ and $\alpha(\upsilon)$ instantiates the alien property in some world). $f$ maps all three intensions to the empty intension. So: $f(\llbracket\lambda\tau.F\upsilon\rrbracket^{\alpha}_{I^{T2}})$ is the empty intension. And $\llbracket\lambda\tau.F\upsilon\rrbracket^{\alpha}_{I^*}$ is also the empty intension. Thus $f(\llbracket\chi_1\rrbracket^{\alpha}_{I^{T2}}) = f(\llbracket\lambda\tau.F\upsilon\rrbracket^{\alpha}_{I^*}) = \llbracket\lambda\tau.F\upsilon\rrbracket^{\alpha}_{I^*} = \llbracket\chi_1\rrbracket^{\alpha}_{I^*}$.
- Suppose 3* holds. Because $\chi_1$ is a 1 place predicate, $\chi_1 \sim_{\beta\eta} F$ and so $\llbracket\chi_1\rrbracket^{\alpha}_{I^{T2}} = \llbracket F\rrbracket^{\alpha}_{I^{T2}}$ and $\llbracket\chi_1\rrbracket^{\alpha}_{I^*} = \llbracket F\rrbracket^{\alpha}_{I^*}$. $\llbracket F\rrbracket^{\alpha}_{I^{T2}}$ is the alien intension $i^a$, which $f$ maps to the empty intension $i^e$: $f(\llbracket F\rrbracket^{\alpha}_{I^{T2}}) = i^e$. And, of course, $\llbracket F\rrbracket^{\alpha}_{I^*} = i^e$. So $f(\llbracket\chi_1\rrbracket^{\alpha}_{I^{T2}}) = \llbracket\chi_1\rrbracket^{\alpha}_{I^*}$.

Given that, $f(\llbracket\chi_1\rrbracket^\alpha_{I^{T2}}) = \llbracket\chi_1\rrbracket^\alpha_{I^*}$, we can see that $\exists\chi_1$ is not diseased:

$$\begin{aligned}
\llbracket\exists\chi_1\rrbracket^\alpha_{I^{T2}} &= \llbracket\exists\rrbracket^\alpha_{I^{T2}}(\llbracket\chi_1\rrbracket^\alpha_{I^{T2}}) \\
&= \llbracket\exists\rrbracket^\alpha_{I^*}(f(\llbracket\chi_1\rrbracket^\alpha_{I^{T2}})) && \text{by the definition of } I^{T2} \\
&= \llbracket\exists\rrbracket^\alpha_{I^*}(\llbracket\chi_1\rrbracket^\alpha_{I^*}) \\
&= \llbracket\exists\chi_1\rrbracket^\alpha_{I^*}
\end{aligned}$$

(iv) Suppose $\zeta$ is $(\lambda\tau_*.\chi)\sigma_1...\sigma_o$. Begin by noting that, by the induction hypothesis, one of 1*-3* holds for $\chi$. Note that the same will hold of $\lambda\tau_*.\chi$. Now consider $(\lambda\tau_*.\chi)\sigma_1...\sigma_o$. We know that one of 1*-3* holds of $(\lambda\tau_*.\chi)$. Let's consider the cases in turn.

- Suppose 1* holds. Then, because $\sigma_1...\sigma_o$ are all undiseased, $(\lambda\tau_*.\chi)\sigma_1...\sigma_o$ is also undiseased.
- Suppose 2* holds. Then $(\lambda\tau_*.\chi)\sigma_1...\sigma_o$ is equivalent to $(\lambda\tau_1....\lambda\tau_o....\lambda\tau_n.Fv)\sigma_1...\sigma_o$. If $v$ is not bound by the first $o$ lambdas, then this is equivalent to $\lambda\tau_{o+1}....\lambda\tau_n.Fv$ and 2* holds of $(\lambda\tau_*.\chi)\sigma_1...\sigma_o$. If $v$ is bound by one of the first $o$ lambdas – say the $p$th lambda – then this is equivalent to $\lambda\tau_{o+1}....\lambda\tau_n.F\sigma_p$. If $\sigma_p$ is a variable, then 2* holds. If $\sigma_p$ is a name, note that $F\sigma_p$ is undiseased (it picks out the empty proposition on either interpretation). So, $\lambda\tau_{o+1}....\lambda\tau_n.F\sigma_p$ is undiseased and 1* holds of $(\lambda\tau_*.\chi)\sigma_1...\sigma_o$.
- Suppose 3* holds, and $\lambda\tau_*.\chi$ is equivalent to $\lambda\tau_1...\lambda\tau_n.F$ and thus $(\lambda\tau_*.\chi)\sigma_1...\sigma_o$ is equivalent to $(\lambda\tau_1....\lambda\tau_n.F)\sigma_1...\sigma_o$. If $o \leq n$ then $(\lambda\tau_*.\chi)\sigma_1...\sigma_o$ is equivalent to $(\lambda\tau_1....\lambda\tau_o....\lambda\tau_n.F)\sigma_1...\sigma_o$, which is equivalent to $\lambda\tau_{o+1}....\lambda\tau_n.F$, and so 3* holds of $(\lambda\tau_*.\chi)\sigma_1...\sigma_o$. If $o = n+1$, then $(\lambda\tau_*.\chi)\sigma_1...\sigma_o$ is equivalent to $(\lambda\tau_1....\lambda\tau_n.F)\sigma_1...\sigma_{n+1}$, which is equivalent to $F\sigma_{n+1}$. If $\sigma_{n+1}$ is a variable, then $F\sigma_{n+1}$ is of the form $Fv$ and 2* holds of $(\lambda\tau_*.\chi)\sigma_1...\sigma_o$. If $\sigma_{n+1}$ is a name, then $F\sigma_{n+1}$ is undiseased (for the reason given in the previous bullet) and 1* holds of $(\lambda\tau_*.\chi)\sigma_1...\sigma_o$. $\square$

**Corollary** (Second Result). *For any sentence $\phi$ and assignment function $\alpha$: $\llbracket\phi\rrbracket^\alpha_{I^{T2}} = \llbracket\phi\rrbracket^\alpha_{I^*}$.*

*Proof.* By our theorem, for any sentence $\phi$, 1*-3* must hold of it. 2* and 3* can't hold of sentences (3* doesn't hold of any formula and 2* only holds of open formulas or $n > 0$ place predicates). So, 1* must hold of $\phi$. $\square$