Can Natural Teleology Ground Logic?

Jeongdam Choi

December 20, 2025

Abstract

This essay examines Dennett's skepticism about original intentionality and his attempt to ground meaning in natural selection and natural teleology. I argue, first, that the indeterminacy problem on which Dennett's skepticism relies does not in itself undermine original intentionality. Second, I argue that the natural-teleological account, when applied to logic, is question-begging because it appeals to the notion of truth, which the Tarskian analysis reveals to presuppose definite meanings for each logical connective. I conclude that our use of logic demonstrates a domain in which the doctrine of original intentionality must be presupposed.

1 Introduction

1.1 The Doctrine of Original Intentionality

Dennett (1981) observes a major disagreement in discussions surrounding the intentionality of the mind. The disagreement is about the doctrine of original intentionality. Dennett puts it as follows:

> The doctrine of original intentionality is the claim that whereas some of our artifacts may have intentionality derived from us, we have original (or intrinsic) intentionality, utterly underived.

Dennett illustrates it with the following example. Consider a vending machine, call it "two-bitser", that accepts U.S. quarters. Allowing ourselves some metaphorical use of language, say that when a two-bitser accepts a quarter, the machine goes into a state Q which "means" *x is a quarter*. Interestingly, vending machines cannot tell Panamanian balboas and U.S. quarters apart. When a two-bitser accepts a balboa, then, should we say that the machine goes into state Q (hence a "misrepresentation"), or does it go into a new state B which "means" *x is a balboa* (hence a "correct representation")?

A straightforward answer is that it depends on the intention of the designer. If an American mechanic designed the two-bitser with the intention of detecting U.S.

quarters and installed it in the U.S., the machine goes into state Q, and a balboa fed to the machine triggers a misrepresentation. Had the designer been Panamanian, the opposite would have been the case. Hence the machine's intentionality is *derived* from that of the designer.

This much is uncontroversial, since a two-bitser is "just an artifact". The debate concerns whether the same is true of humans. It is held *prima facie* that when Jones sees a horse and utters "horse", he is in a state H, meaning *x is a horse*. One day, Jones is confronted with a *schmorse*, a creature that resembles a horse yet is taxonomically not a horse. Does this encounter cause Jones to be in state H, or does it cause Jones to be in a new state SH meaning *x is a schmorse*? Or could it have been that it was not state H, but rather state H*, meaning *x is a horse or a schmorse or ...* that Jones has been in all along? Here is where Dennett draws the battle line:

> *However hard it may be to determine exactly which state he is in, he is really in one or the other...* Anyone who finds this intuition irresistible believes in original intentionality, and has some distinguished company: Fodor, Searle, Dretske, Burge, and Kripke... Anyone who finds this intuition dubious if not downright dismissible can join me, the Churchlands, Davidson, Haugeland, Millikan, Rorty, Stalnaker, and our distinguished predecessors, Quine and Sellars, in the other corner. (My emphasis)

1.2 Dennett's Mother Nature

Dennett claims that no fact about the individual can solve the indeterminacy problem regarding their mental states. What is needed to resolve the indeterminacy is to recognize that our intentionality is just as derived as that of a two-bitser, except that the designer is not human, but "Mother Nature".

To illustrate, it is well-known that frogs "mistakenly" snap at lead pellets for flies. But what justifies us in saying that snapping-at-lead is a "mistake"? For, if the state associated with the frog's snapping means *x is a fly*, then snapping-at-lead is a mistake; but had the state meant *x is a fly or a lead pellet*, then snapping-at-lead would not be a mistake. So what justifies us in saying that the frog's snapping means the former and not the latter?

Dennett argues that the justification is not to be found in the frog's neurophysiology, but in the evolutionary history guided by natural selection that led frogs to snap at things. It then becomes evident that the frog's snapping has been "selected for" its function to obtain nutrients, so a frog snapping at a lead pellet is a mistake. Thus, just like a two-bitser, the intentionality of a frog is derived from its designer; the designer being "Mother Nature".

And if frogs have derived intentionality, the same would be true for more complex organisms, including ourselves. Thus, a Dennettian may argue that Jones referring to

a schmorse as "horse" is a mistake, *if* (but not necessarily *only if*)

1. Our disposition to utter "horse" upon seeing horses was "selected for" its ability to recognize tamable and ridable animals.

2. Schmorses are either not tamable or not ridable.

Following Fodor (1996), we can "deconstruct" Dennett's position as composed of three parts: (1) adaptationism is the true account of natural selection, (2) adaptationism grounds natural teleology (the way of saying that nature "selects" an organ "for" a purpose) — that is, adaptationism provides truth conditions for the propositions of natural teleology, and (3) natural teleology grounds intentionality and meaning.

In this essay, I will begin by simply accepting (1) and (2). So for example, I accept that although it is presumably a nomological necessity that hearts which pump blood are coextensional with hearts which make noise, it is true that hearts have been "selected for" their blood-pumping capability and not for their noise-making capability. The concern of this essay is whether this charitable reading of natural teleology is sufficient to ground intentionality.

2 Recasting the Problem

2.1 Is the Indeterminacy Problem Well-Posed?

Since Dennett cites the indeterminacy problem as a major reason to doubt original intentionality, it will aid our discussion to elucidate the exact structure of the problem as follows.

1. Intentional realists (proponents of original intentionality) hold that:

   (a) There is a class of mental states called *representational mental states*.

   (b) Each representational mental state has a content that represents the world as being a certain way.

   (c) If an agent is in a representational mental state, there should be no indeterminacy as to which representational mental state they are in.

2. From (1b) it follows that:

   (a) Given a representational mental state, there should be no indeterminacy as to whether it represents a world correctly.

3. From (1c) and (2a) it follows that intentional realists hold that:

   (a) If an agent is in a representational mental state, there should be no indeterminacy as to whether their state represents a world correctly.

4. Hence, if there is a counterexample to (3a), intentional realism is false.

Dennett cites various counterexamples to (3a), including the aforementioned horse-schmorse case, but also Putnam's H_2O-XYZ case, and his own *glug* case (Dennett, 1981). However, I argue that the indeterminacy problem as stated above is misformulated, since counterexamples to (3a) will always exist even if intentional realism were true. Granting (1a), (1b), and (1c), say that when Jones utters "chair", he is in a representational mental state whose content is *x is a chair*. Consider a Sorites sequence $(s_1, \ldots, s_n)$ where s_1 is a chair and s_n is a pile of sawdust. (3a) demands that there be a principled criterion as to whether Jones correctly represents s_i for $1 \leq i \leq n$ when he utters "chair". But there isn't an answer to be found, hence (3a) is falsified.

Yet this cannot be a valid argument against intentional realism. At most, it reveals that the content associated with "chair" has a *vague* extension. The fallacy lies in deducing (2a) from (1b).[1] A content can represent the world *as being a certain way* without a predetermined criterion as to which worlds *are correct*. The conclusion to be drawn is not that the meaning is indeterminate, but that a determinate meaning does not necessarily come equipped with a determinate criterion.

That said, the extent to which this point applies to Dennett's examples is not straightforward. Can the vagueness concerning one's uttering of "chair" to half-broken chairs be identified with that of "horse" to schmorses? Irrealists about biological species may be inclined to think so, while essentialists will argue that it is a desperate move to put the two on the same line.[2] We could decide to go on and try to settle this issue, but this would be to burden the discussion with undesirable metaphysical baggage.

My suggestion therefore is that we set aside the question of whether Dennett's examples of indeterminacy genuinely threaten intentional realism. I find that a better way to proceed is to dispense with using physical objects — "natural kind" candidates — to pose the indeterminacy problem, and instead to restrict the problem to languages whose meanings, if they exist, cannot possibly be vague. This shifts the focus to logic.

2.2 Indeterminacy in Logic and Mathematics

A famous skeptical paradox by Kripke (1982) asks whether there is a fact as to whether a speaker, who has never added numbers higher than 50, meant addition, or *quaddition*

[1] It seems to be an *a priori* truth that there cannot be indeterminacy in (1b), i.e., in the ascription of content to mental states, granted they exist. For it should be content, if anything, that constitutes the essential property of a mental state, and hence the basis for rigidly designating it. See Kripke (1980), pp. 146–155, for an expression of a similar view.

[2] To put the matter more generally, it is held by some that "natural kinds" are categorically distinct; there cannot be a smooth transition from one kind to another (Ellis, 2001). Under this assumption, indeterminacy problems restricted to natural kinds may allow inferring (2a) from (1b). Though I still think there remains a difficulty, since the premise required is that there is *no* candidate for Jones' mental state whose content is vague. This is much stronger than the statement that *not all* candidates for Jones' mental state have vague content.

(defined below), with '+'.

$$x \oplus y = \begin{cases} x + y & (x, y < 50) \\ 5 & (\text{otherwise}) \end{cases}$$

For reasons that will become clear shortly, I will operate with a slight variation of Kripke's case, applied to logic instead of arithmetic. Consider the material implication symbol '$\rightarrow$'.[3] Let *quimplication* be the operation that coincides with implication if it appears in a sentence with less than 50 tokens, and evaluates to false otherwise. For example, "$p \rightarrow p$" is true, but "$(p_1 \wedge p_2 \wedge \cdots \wedge p_{50}) \rightarrow p_1$" is false.

Suppose that Alice has, in her whole lifetime, dealt with a sentence of more than 50 tokens only once — for more drama we may consider her dead, so that no further data about her use of language are obtainable. Let us further assume that the said sentence, say σ, is true, but Alice had evaluated it to be false. It turns out that σ is indeed false if '$\rightarrow$' is taken to stand for quimplication. We are now faced with an indeterminacy akin to the horse-schmorse case — a case I claim to be better suited for discussing intentionality, for now the indeterminacy cannot be imputed to the vagueness of content boundary.[4] So let us recast the indeterminacy problem of intentionality as follows. *Did Alice mean implication with '$\rightarrow$' and make a mistake, or did Alice mean quimplication with '$\rightarrow$'?*

3 The Case for Logic

3.1 The Dennettian View of Logic

A naive Dennettian would take the statements "Alice means implication with '$\rightarrow$'" and "Alice means quimplication with '$\rightarrow$'" to be *equally* justified. Indeed, they both satisfy the criteria for "real patterns" laid out by Dennett (1991). Both describe Alice's use of '$\rightarrow$' as patterns—*inference rules*—in her behavior. For instance, they express observations such as "when Alice approves 'p' and '$p \rightarrow q$', she is disposed to approve 'q'". The difference is that the latter requires the number of tokens in '$p \rightarrow q$' to be less than 50. Nevertheless, the two patterns are both *real*, in the sense that, if Jones uses the first pattern and Brown the second to predict Alice's uses of '$\rightarrow$' and bet accordingly, "they will both get rich."[5]

[3] Although for convenience I am using symbols of formal logic for illustration, the symbols should be broadly construed as encompassing natural language. So '$\rightarrow$' also stands for the English words "implies" and "if", while "$p \rightarrow q$" also stands for the English sentence "If p then q".

[4] This point is also emphasized by Kripke (1982): "The point is... that anything in my head leaves it undetermined what function 'plus' (as I use it) denotes... The skeptical problem indicates no vagueness in the concept of addition (in the way there is vagueness in the concept of greenness)... The skeptical point is something else."

[5] Indeed, Brown will be *richer* than Jones, since only he would have correctly predicted Alice's response to σ.

But a more sophisticated Dennettian may resist succumbing to the indeterminacy so easily. Instead, she could attempt to justify the implication interpretation over the quimplication interpretation by appeal to natural teleology. Previously we have seen how natural teleology licenses us to say that a frog snapping at a lead pellet is a mistake, because the frog's ability to snap at things has been "selected for" its function of obtaining nutrients. That is, natural teleology justifies "A frog's snapping means fly-catching" over "A frog's snapping means fly-or-pellet catching."

There are two ways in which a Dennettian can extend this reasoning to the realm of logic so as to preclude the quimplication interpretation in favor of the implication interpretation. I will call these the robust reasoning and the trivial reasoning.

- The *robust* reasoning defines truth independently of evolutionary considerations and, from that definition of truth, claims that:

  1. Truth has utility.
  2. Therefore, truth-preserving inference rules are more evolutionarily advantageous than non-truth-preserving inference rules.
  3. Inferences which take '$\rightarrow$' as quimplication are not truth-preserving, whereas inferences which take '$\rightarrow$' as implication are.

  Therefore, Alice means implication by '$\rightarrow$'.

- The *trivial* reasoning bypasses the notion of truth and claims directly that some inference rules are more evolutionarily advantageous than others. It then follows that Alice's inference rules will converge upon the rules that are most evolutionarily advantageous, which happen to take '$\rightarrow$' as implication.

However, I argue that both reasonings are problematic. I argue that the robust reasoning is regressive, because defining truth presupposes prior facts about what Alice means by her logical constants. I argue that the trivial reasoning is irrelevant, because it fails to capture one of the core components of natural teleology, namely intensionality (with an 's').

3.2 The Robust Reasoning

Let us begin with the robust reasoning. It is clear that the robust reasoning cannot define truth with the naive correspondence theory, i.e. a sentence is true iff what it represents is the case. For to operate with such a notion of truth would be to admit the existence of mental representations, which Dennettians flatly reject. The robust reasoning needs to define truth without reference to mental concepts.

Incidentally, one of the leading motivations for Tarski in developing his theory of truth was to fulfill this aim. Tarski was a moderate physicalist, meaning that he

believed all scientific facts, claims, and concepts to be reducible to those of physics and logic (construed to include set theory). He wanted a notion of truth compatible with his physicalism; the end result was his theory of truth.

Following Field (1972), we can understand Tarski's theory of truth as composed of two steps. The first step enumerates the "primitive denotations" of the target language. Here is a toy example. Let L be a language with three names $\ulcorner m \urcorner, \ulcorner v \urcorner, \ulcorner j \urcorner$, two one-place predicate symbols $\ulcorner \text{Terr} \urcorner, \ulcorner \text{Gas} \urcorner$, and two logical constants $\dot{\rightarrow}, \dot{\neg}$.[6] The primitive denotations for L are:

> A name n *refers* to an object o iff
>
> $$(n, o) \in \{(\ulcorner m \urcorner, \text{Mercury}), (\ulcorner v \urcorner, \text{Venus}), (\ulcorner j \urcorner, \text{Jupiter})\}.$$
>
> A predicate P *applies* to an object o iff
>
> $$(P, o) \in \{(\ulcorner \text{Terr} \urcorner, \text{Mercury}), (\ulcorner \text{Terr} \urcorner, \text{Venus}), (\ulcorner \text{Gas} \urcorner, \text{Jupiter})\}.$$

The second step provides the recursive definition of truth based on the primitive denotations.

> Define the *truth-in-L* predicate T_L, ranging over the sentences of L, as follows:
>
> - $T_L(\ulcorner Pn \urcorner)$ iff P applies to the object referred to by n.
> - $T_L(\ulcorner A \dot{\rightarrow} B \urcorner)$ iff $T_L(\ulcorner A \urcorner)$ implies $T_L(\ulcorner B \urcorner)$.
> - $T_L(\ulcorner \dot{\neg} A \urcorner)$ iff it is not the case that $T_L(\ulcorner A \urcorner)$.

Although the significance of Tarski's truth predicate is undoubted in mathematics, Field (1972) pointed out its limitation for use in the philosophical investigation of natural languages. Specifically, he pointed out that the Tarskian definition takes the primitive denotations of the target language as a brute fact. This leads to the following problem. Whether n refers to o depends on facts about the speaker and/or the world (be they mental, physical, communal, naturalistic, behavioral, etc.). Therefore, had the facts about the speaker and/or the world been different, $\ulcorner m \urcorner$ could have referred to Jupiter, and $\ulcorner \text{Gas}(m) \urcorner$ could have been true (in L). However, this counterfactual does not hold in the present setting, as there is nothing concerning the facts about the speaker or the world in the Tarskian definition.

Field attempted to overcome this problem and complete Tarski's physicalist program by providing a generic reduction of primitive denotations. Specifically, he aimed to fill in the blanks for the following clauses with physical facts about the speaker and/or the world.

[6] The square quotes and dots serve to distinguish the object language from the metalanguage.

n as used by the speaker S refers to o iff _____.
P as used by the speaker S applies to o iff _____.

Field's original idea was to fill in the blanks with causal theories of reference of the sort proposed by Kripke. However, numerous arguments against causal theories as necessary and sufficient conditions for reference have since been raised (see Adams and Aizawa (2021), section 4), which eventually led Field to abandon his project.

It thus seems that this is the point from which the robust reasoning should pick up. The prospect seems optimistic at first glance, for Dennettians need not be fixated on causal theories. Dennettians can instead supplement the first step of the Tarskian definition, namely primitive denotations, with natural-teleological accounts. The second step would then go through, resulting in a robust definition of truth.

However, Soames (1984) points out that Fodor underestimated the force of his own criticism of the Tarskian definition. Just as whether $\ulcorner m \urcorner$ refers to Mercury depends on facts about the speaker, so too does whether $\dot{\rightarrow}$ refers to implication depend on facts about the speaker. Yet the following clause in the Tarskian definition

$T_L(\ulcorner A \dot{\rightarrow} B \urcorner)$ iff $T_L(\ulcorner A \urcorner)$ implies $T_L(\ulcorner B \urcorner)$,

takes as a brute fact that the speaker uses $\dot{\rightarrow}$ to refer to implication. A generic definition of truth should therefore replace the above clause with the following set of clauses, the blank of which should be filled by facts about the speaker and/or the world.

- $\dot{\rightarrow}$ as used by S refers to a logical operator ξ iff _____.
- $T_L(\ulcorner A \dot{\rightarrow} B \urcorner)$ iff $\dot{\rightarrow}$ as used by S refers to ξ and $\xi(T_L(\ulcorner A \urcorner), T_L(\ulcorner B \urcorner))$.

Soames uses this point to challenge the physicalist program envisioned by Field. However, this point poses an even more serious problem for the robust reasoning, since the first clause is precisely what the robust reasoning set out to explain. In order to decide whether a speaker means implication or quimplication by '$\rightarrow$', the robust reasoning proposed first to define truth independently of evolutionary considerations and then to favor implication over quimplication on the grounds of its truth-preservingness. Yet it turns out that such a definition of truth itself requires a prior specification of how each logical constant, as used by the speaker, is to be interpreted, thereby resulting in a self-regress. Whether it is inferences that take '$\rightarrow$' as implication, or inferences that take '$\rightarrow$' as quimplication, that are truth-preserving viciously depends on whether the speaker means implication or quimplication with '$\rightarrow$'.

3.3 The Trivial Reasoning

Since the robust reasoning is bust, let us look into the trivial reasoning. To reiterate, the trivial reasoning bypasses the notion of truth and claims directly that some

inference rules are more evolutionarily advantageous than others.

The initial plausibility of this reasoning is suggested by the following thought experiment. Consider three linguistic communities C_1, C_2, C_3. Members of C_1 accept as a valid inference rule *mighty ponens*, which deduces any conclusion q from any premise p. Clearly, C_1 will not fare well. It will go extinct rather quickly, for its members may at any moment "infer" that killing oneself is desirable. Meanwhile, members of C_2 accept *feeble ponens*, which deduces p from p, as the only valid inference rule. Although C_2 may fare better than C_1, this is not saying much. The information its members can utilize is limited to their immediate sense data, and with such austerity, they too will face their demise sooner or later. Finally, members of C_3 accept *modus ponens*. That is, a certain symbol '$\rightarrow$' is used by its members in the following way: if one accepts p and $p \rightarrow q$, then one is disposed to accept q. As evidenced by our evolutionary success, this inference rule hits the sweet spot between the too-strong *mighty ponens* and the too-weak *feeble ponens*. The upshot is that this thought experiment, crude as it may be, justifies talk of some inference rules being more evolutionarily advantageous than others.

The problem with the trivial reasoning is that this observation is irrelevant to the theory of natural teleology. Natural teleology is the way of saying that nature "selects" an organ (or "develops" dispositions) "for" a purpose. A central tenet of natural teleology is that the "for" here is *intensional*. Thus, even if hearts that make noise and hearts that pump blood are coextensional, "Hearts were selected for noise-making" is false, while "Hearts were selected for blood-pumping" is true. This is important for Dennett, because Dennett takes intensionality to be "a defining mark of intentionality" (Dennett, 1981). One can believe a celestial body to be the morning star without believing it to be the evening star. Thus, to achieve his aim of grounding meaning and intentionality in natural selection, he needs to give an account of how intensionality appears in the natural-teleology picture. Here is how he gives such an account:

> The disappearance of intensionality at the macromolecular level at first seems a telling objection to the persistent use of intentional idioms to characterize that level, but if we leave it at that we miss a still deeper level at which the missing intensionality re-appears. The synthetase may not desire that isoleucine-AMP be the intermediate amino acid, but it is only *qua* intermediate that the isoleucine is "desired" at all... And while the proof-reading enzyme has no inkling that it is correcting errors *qua* errors, Mother Nature does! That is, it is only *qua* error that the items thus eliminated provoked the creation of the "proof-reading" competence of the enzymes in the first place. The enzyme itself is just one of Nature's lowly soldiers, "theirs not to reason why, theirs but to do or die," but *there is* a reason why they do what they do, a reason "recognized" by natural

selection itself...

> How can natural selection do this without intelligence? It does not consciously seek out these rationales, but when it stumbles on them, the brute requirements of replication ensure that it "recognizes" their value. The illusion of intelligence is created because of our limited perspective on the process; evolution may well have tried all the "stupid moves" in addition to the "smart moves", but the stupid moves, being failures, disappeared from view. All we see is the unbroken string of triumphs. When we set ourselves the task of explaining why *those* were the triumphs, we uncover the reasons for things—the reasons already "acknowledged" by the relative success of organisms endowed with those things.

But even if we grant that hearts were selected for blood-pumping and that isoleucine was selected *qua* intermediate amino acid, this falls short of justifying that our use of '$\rightarrow$' was selected (developed) *qua* implication. As emphasized in the quoted passage, the reason we can say that hearts were selected for blood-pumping is because it makes sense to take blood-pumping as an *explanation* for the evolutionary success of organisms with hearts. To apply the same reasoning to logic, it should make sense to take *modus ponens* as an explanation for the evolutionary success of our way of using '$\rightarrow$'. However, the trivial reasoning falls short of making such sense. The trivial reasoning is simply an observation of the following facts:

1. The way we use '$\rightarrow$' must be evolutionarily advantageous, for we are the end product of evolution.

2. *Modus ponens* is coextensional with the way we use '$\rightarrow$'.

Thus, the trivial reasoning only concerns the extension of *modus ponens*. It thus allows substitution for any inference rule extensionally equivalent to *modus ponens*. For example, let *skidus ponens* be the inference rule that deduces q from p and $p \rightarrow q$, *if snow is white*. Say that p *skimplies* q iff p implies q and snow is white. It follows that if *modus ponens* were an explanation for our successful use of '$\rightarrow$' as implication, so would *skidus ponens* be an explanation for our successful use of '$\rightarrow$' as skimplication, at least according to the trivial reasoning.

To put it shortly, the trivial reasoning is *too* trivial to satisfy the intensionality of natural teleology. In order to fill in the missing intensionality, a metalogical concept such as truth is required. We would then be able to say that our use of logical constants was *selected for* truth preservation. This would render the skimplication explanation incoherent, since whether snow is white serves no purpose in the explanation *qua* truth preservation. However, this is to fall back to the robust reasoning, which we have concluded to be regressive.

3.4 Is the Indeterminacy of Logic Tolerable?

The remaining choice for Dennettians is to embrace the indeterminacy in interpreting Alice's logical constants in its fullest sense. Since this would equally apply to everyone else, it follows that logic as a whole is neither absolute nor normative, but rather is just one set of descriptions about patterns in thinking to which humans in general tend to conform.

However, a well-developed line of thought, prominently featured in the 19th- and 20th-century German-speaking world, asserts that such a view of logic is a *non sequitur*. The line was developed as a reaction against the then-fashionable psychologism, which holds that since psychology is the empirical science that explains why people think so-and-so, and logic is about regularities in thought, it follows that logic is subsumed under psychology.

Although it is tempting to identify Dennett's view with psychologism, with "psychology" construed to include neurophysiology and behaviorism, this accusation in itself would have been unjust, for Dennett is also equipped with the natural teleology story that could have secured some sense of normativity or correctness. However, now that we have seen that this story is ineffective for logic, it seems that the Dennettian view really is committed to a full-fledged psychologism, and is prone to the difficulties raised against it.

But instead of investigating these difficulties, which others have done and continue to do so to a great extent (see Kusch (2024)), it would be more appropriate in this essay to point out that even if the difficulties of psychologism could somehow be overcome, there are further reasons to doubt that it could be consistent *vis-à-vis* the Dennettian view. For, what is unique to the Dennettian view is that it tries to stay as an intermediate position between eliminativism and realism about mental contents. Dennett distances himself from eliminativism by appreciating the power of folk-psychology and holding that attributions of meanings or beliefs are justified, while also distancing from realism by claiming that there is no answer as to what one *truly* means or believes. Thus he is led to claims such as "there could be two [disagreeing] systems of belief attribution to an individual... and yet where no deeper fact of the matter could establish that one was a description of the individual's *real* beliefs and the other not." (Dennett, 1991)

Frege (1884) famously raised the point that the notion of disagreement only makes sense against a common framework of logic. To illustrate how Frege's point is relevant, consider a scenario where Jones and Brown attribute to Alice two different systems of belief about logic. Jones attributes the standard beliefs such as "From 'p' and '$p \to q$' one ought to deduce 'q'", while Brown attributes bizarre beliefs that nonetheless explain Alice's behavior equally well. Frege would ask: how can it be substantiated that the systems attributed by Brown and Jones are truly different? Perhaps Brown and Jones write out what they take "truth-in-L_A", where L_A is Alice's language, to

be as follows.

- Jones claims: "$T_A(\ulcorner \phi \dot{\rightarrow} \psi \urcorner)$ if and only if $T_A(\ulcorner \phi \urcorner)$ implies $T_A(\ulcorner \psi \urcorner)$"
- Brown claims: "$T_A(\ulcorner \phi \dot{\rightarrow} \psi \urcorner)$ if and only if $T_A(\ulcorner \phi \urcorner)$ implies $T_A(\ulcorner \psi \urcorner)$, and snow is white."

However, this falls short of showing conclusively that their attributions are different, since it could be that Jones happens to mean *skimplication* with "implies", rendering the two attributions equivalent. That is, writing T_J for "truth-in-L_J" where L_J stands for Jones' language, it could turn out that:

$T_J(\ulcorner \phi \dot{\rightarrow} \psi \urcorner)$ if and only if $T_J(\ulcorner \phi \urcorner)$ implies $T_J(\ulcorner \psi \urcorner)$, and snow is white.

To escape from such regress, a shared framework of logic must be firmly in place. Only when it is given *simpliciter* that Brown and Jones both mean implication with "implies" can we legitimately say that they attribute to Alice different systems of belief about logic. But then the *simpliciter* phrase refutes the very skepticism that entertained this scenario, namely, skepticism about the doctrine of original intentionality.

In conclusion, the Dennettian view, when subjected to logic, faces a dilemma. It cannot furnish a natural-teleological explanation since it is prohibited from appealing to the utility of truth. Yet it cannot simply embrace the indeterminacy, since it would then be unintelligible that there be *differing* attributions of beliefs. In both cases we are led to conclude that our mental states for our logical constants do have definite meanings.

4 Conclusion

In *Darwin's Dangerous Idea*, Dennett (1995) elaborates his project to show that intentionality can be explained entirely by "cranes" — the mindless process of natural selection — without reliance on "skyhooks," namely, the doctrine of original intentionality. I have argued, however, that this picture overlooks the very "bedrock" that enables cranes to stand in the first place, the bedrock being logic.

Readers may wonder where this leaves us with respect to logic. One might wonder whether a stronger theory could ultimately explain away our most basic intentionality. Pursuing that question lies beyond the scope of this essay. My aim has been more modest: to illustrate problems in natural-teleological explanations for logic. I therefore close here, but with a brief remark that I align myself with a line of thought found in Kant, Frege, Wittgenstein, and Putnam — namely, that logic is not something that requires explanation, but is presupposed in every act of explanation. (See Conant (1992) for a lucid survey of this thought.)

References

Adams, F. and K. Aizawa (2021). Causal Theories of Mental Content. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021 ed.). Metaphysics Research Lab, Stanford University.

Conant, J. (1992). The Search for Logically Alien Thought. *Philosophical Topics 20*(1), 115–180.

Dennett, D. C. (1981). Evolution, Error, and Intentionality. In *The Intentional Stance*, pp. 287–321. Cambridge, MA: MIT Press.

Dennett, D. C. (1991). Real Patterns. *The Journal of Philosophy 88*(1), 27–51.

Dennett, D. C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon & Schuster.

Ellis, B. (2001). *Scientific Essentialism*. Cambridge Studies in Philosophy. Cambridge: Cambridge University Press.

Field, H. (1972). Tarski's Theory of Truth. *The Journal of Philosophy 69*(13), 347–375.

Fodor, J. A. (1996). Deconstructing Dennett's Darwin. *Mind & Language 11*(3), 246–262.

Frege, G. (1884). *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*. Evanston, IL: Northwestern University Press. Original work published 1884. Translated by J. L. Austin.

Kripke, S. A. (1980). *Naming and Necessity: Lectures Given to the Princeton University Philosophy Colloquium*. Cambridge, MA: Harvard University Press.

Kripke, S. A. (1982). *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Cambridge, MA: Harvard University Press.

Kusch, M. (2024). Psychologism. In E. N. Zalta and U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2024 ed.). Metaphysics Research Lab, Stanford University.

Soames, S. (1984). What Is a Theory of Truth? *The Journal of Philosophy 81*(8), 411–429.