Can Natural Teleology Ground Logic?

Jeongdam Choi

September 14, 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; and second, that the natural-teleological account, when applied to logic, is question-begging because it appeals to the utility of truth. 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 why is it a "mistake"? The answer is not 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 too, including ourselves.

Following Fodor (1996), we can more precisely "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 —, and (3) natural teleology grounds intentionality. 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 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 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

[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.

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.[2]

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 view so, while essentialists will argue that it is a desperate move to put the two on the same line.[3] We could decide to go on and try to settle this issue, but this would be to laden 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 meaning, if they exist, cannot *possibly* be vague. This leads us to shift 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* (defined below), with '+'.

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

For reasons to come, I will operate with a slight variation of Kripke's

[2]If it helps one may think of such meaning as a decision tree allowing supervaluationistic truth value. The meaning — the decision tree — is quite determinate; it simply has fuzzy values in some of its nodes.

[3]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 is difficulty remaining, 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.

case, applied to logic instead of arithmetic. Consider the material implication symbol '$\rightarrow$'.[4] 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. Did Alice mean implication with '$\rightarrow$' but had made a mistake, or did Alice mean quimplication with '$\rightarrow$'?

I claim that the case for '$\rightarrow$' is better suited for discussing intentionality, for now the indeterminacy cannot be imputed to the vagueness of content boundary.[5] So let us recast the indeterminacy problem of intentionality as follows. *When we use logical vocabulary such as "implies", are we in mental states with determinate meanings?* I will now outline how a Dennettian might respond to this question, and then argue that the responses are question-begging. Hence the doctrine of original intentionality must hold true at least for our use of logic.

3 The Case for Logic

3.1 The Dennettian View of Logic

Presumably, a Dennettian would hold the statement "Alice means implication with '$\rightarrow$'" and "Alice means quimplication with '$\rightarrow$'" to be both justified. Indeed they both satisfy the criteria for "real patterns" laid out

[4] 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". And for this reason, I will from here on use the expression "logical vocabulary" instead of "logical symbols".

[5] 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."

by Dennett (1991). Both describe Alice's use of '$\rightarrow$' as a pattern in her behavior, consisting of statements 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. This difference notwithstanding, the two patterns are both *real*, in the sense that had Jones used the first pattern and Brown the second to predict Alice's future uses of '$\rightarrow$' and bet accordingly, "they will both get rich". (Indeed, Brown would have gotten *richer* than Jones, since only he would have correctly predicted Alice's response for σ.)

But a more ambitious Dennettian may not succumb to the indeterminacy so easily. Instead, she could attempt to justify the implication interpretation over quimplication through natural teleology. Previously we have seen how natural teleology paves the way to say that a frog snapping at a lead pellet is a mistake, because the frog's ability to snap at things had been "selected for" its function to obtain nutrients. A Dennettian could argue likewise for logic: our ability to reason with logic had been "selected for" its function to deduce true sentences from true premises, because knowing more truths than falsities is evolutionarily beneficial. It then follows that since deductions with '$\rightarrow$' as quimplication are vastly inconsistent, Alice meant implication with '$\rightarrow$', and her deducing σ to be true was a mistake.

3.2 Logic, Truth, and Utility

However, I argue that this line of response has issues, due to its appeal to the notion of truth. It is a well-received moral from the works of Tarski that the notion of truth is relative to language. Hence, when a Dennettian claims that Alice's knowing true sentences is beneficial to Alice, the "true" here really means "true-in-L_A", where L_A stands for Alice's language. Importantly, Soames (1984) points out that the scope of a language includes its logical vocabulary. Hence, the clauses for Tarskian truth definition, such as:

$$T_A(\ulcorner \phi \dot{\rightarrow} \psi \urcorner) \text{ if and only if } T_A(\ulcorner \phi \urcorner) \text{ implies } T_A(\ulcorner \psi \urcorner) \ ^{6}$$

where T_A stands for "truth-in-L_A" presumes that '$\rightarrow$' *means* implication in L_A. Without this premise, the definition of truth cannot even get off the

[6]The dot above $\rightarrow$ indicates that the symbol is mentioned, not used.

ground. Soames uses this point to argue against a physicalist reduction of truth envisioned by Field (1972), but we can also use his point to deliver an attack against the natural-teleological explanation for logic. The natural-teleological explanation appeals to the utility of truth, specifically "truth-in-L_A", to decide what the (derived) meanings of the logical vocabulary as used by Alice are. However, the very definition of "truth-in-L_A" is dependent on the said meaning, thus begging the question.

Yet Dennettians need to appeal to *some* notion of truth, for it is the utility of truth, if anything, that endows a certain class of sentences and deductions to be more "evolutionarily beneficial" than others.[7] That said, one possible move for Dennettians is to simply define "truth-in-L_A" as the class of sentences that, were Alice to believe them, would be useful to her. This would be to adopt the pragmatist theory of truth, and the claim that truth is evolutionarily beneficial would follow *tout court*. However, not only would it then be utterly unclear as to why implication is more truth-preserving (i.e. utility-preserving) than quimplication, it is even unclear if implication really *is* more truth-preserving than quimplication — not to mention the more serious problems of the pragmatist theory that made it fall out of fashion.

Can Dennettians choose to simply bypass giving a substantial account of truth by adopting a deflationist view? Again, the verdict lies on whether deflationism can explain the utility of truth. Burgess and Burgess (2011) gives a sketch of such an explanation. Call a belief *directly action-guiding* if having that belief leads the agent to do a certain action unless the agent is weak-willed, e.g.

> (2) Going to the gym now will lead to the outcome most preferred.

Say a belief is useful if it is useful in obtaining the agent's most preferred outcome. It follows tautologically that (2) is a useful belief, if, going to the gym now will lead to the outcome most preferred, which in the deflationist

[7]Dennett (1991) also notes the point, writing that "Even someone who... has seen the futility of correspondence theories of truth must accept the fact that within the natural ontological attitude we sometimes explain success by correspondence: one does better navigating off the coast of Maine when one uses an up-to-date nautical chart than one does when one uses a road map of Kansas. Why? Because the former accurately represents the hazards, markers, depths, and coastlines of the Maine coast, and the latter does not."

view of truth is equivalent to saying that (2) is true. So deflationists can explain the utility of truth, at least for directly action-guiding beliefs.

Burgess then extends this approach to the general case. The picture is that "since directly action-guiding beliefs will generally be inferred as conclusions in some manner from other beliefs taken as premises, it will be indirectly useful if the other beliefs involved are true and the manner of inferring conclusions from premises involved is truth-preserving". The picture may be more fully articulated as follows. If Alice's belief in

> (3) If I don't have any urgent deadlines today, then going to the gym now will lead to the outcome most preferred.
>
> (4) I don't have any urgent deadlines today.

leads her to believe (2), it follows that:

- (3) and (4) are indirectly useful beliefs, if (2) is a useful belief.
- (2) is a useful belief, if going to the gym now will lead to the outcome most preferred to Alice.
- Going to the gym now will lead to the outcome most preferred to Alice, if it is the case that (a) if Alice doesn't have any urgent deadlines today, then going to the gym now will lead to the outcome most preferred to Alice, and (b) Alice doesn't have any urgent deadlines today.
- (a) is equivalent to (3) being true for Alice, and (b) is equivalent to (4) being true for Alice.
- Hence, (3) and (4) are indirectly useful beliefs if (3) and (4) are true for Alice.

Although this picture may be satisfactory for deflationism *per se*, it poses problems for Dennettians, because this picture once again tacitly depends on the Tarskian approach to truth. Namely, in claiming that (3) being true for Alice is equivalent to (a), it is presumed that Alice means *if* with "if". Had Alice meant *quif* with "if", the truth (specifically, truth-in-L_A) of (3) would not be equivalent to (a), but to

(c) *Quif* Alice doesn't have any urgent deadlines today, then going to the gym now will lead to the outcome most preferred to Alice.

which, even in conjunction with (b), does not necessarily imply that going to the gym now will lead to the outcome most preferred to Alice.

In conclusion, any natural-teleological story that appeals to the utility of truth to decide what an agent means with their logical vocabulary presupposes, via its Tarskian underpinnings, prior facts about the said meanings, begging the question. Hence, even if Dennettians could explain *some* notion of intentionality and correctness by appealing to natural teleology, logic cannot be one of them.

3.3 Is the Indeterminacy of Logic Tolerable?

The remaining choice for Dennettians is to embrace the indeterminacy in interpreting Alice's logical vocabulary 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 19-20th century German-speaking world, asserts that such a view of logic is *non sequitur*. The line was developed as a reaction against 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 thoughts, 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 implicated 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 great extent (see Kusch (2024)), it would be more appropriate in this essay to point out that even if the difficulties of psychol-

ogism 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 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" to be as follows.

- Jones claims: "$T_A(\ulcorner \phi \dot{\to} \psi \urcorner)$ if and only if $T_A(\ulcorner \phi \urcorner)$ implies $T_A(\ulcorner \psi \urcorner)$"
- Brown claims: "$T_A(\ulcorner \phi \dot{\to} \psi \urcorner)$ 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:

> $T_J(\ulcorner \phi \dot{\to} \psi \urcorner)$ 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 the conclusion forced upon is that our mental states for our logical vocabulary 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.

The 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

Burgess, A. G. and J. P. Burgess (2011). *Truth*. Princeton, NJ: Princeton University Press.

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.

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. *Philosophy 56*(217), 431–433.

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.