

# Defending Non-Derived Content

Ken Aizawa and Fred Adams

*In “The Myth of Original Intentionality,” Daniel Dennett appears to want to argue for four claims involving the familiar distinction between original (or underived) and derived intentionality.*

1. *Humans lack original intentionality.*
2. *Humans have derived intentionality only.*
3. *There is no distinction between original and derived intentionality.*
4. *There is no such thing as original intentionality.*

*We argue that Dennett’s discussion fails to secure any of these conclusions for the contents of thoughts.*

*Keywords: Original Intentionality; Derived Intentionality; Daniel Dennett*

In “The Myth of Original Intentionality” (1990), Daniel Dennett appears to want to argue for four claims involving the familiar distinction between original (or underived) and derived intentionality:

1. Humans lack original intentionality.
2. Humans have derived intentionality only.
3. There is no distinction between original and derived intentionality.
4. There is no such thing as original intentionality.

Contrary to the billing we have been given about this paper, we think that Dennett’s discussion fails to secure any of these conclusions for the contents of thoughts.

Some paradigm cases of items bearing derived content are traffic lights, gas gauges, and flags. The paradigm cases of items bearing non-derived content are thoughts, experiences, and perceptions.<sup>1</sup> Roughly speaking, the idea is that derived content arises from the way in which items are handled or treated by intentional agents. For the most part, things with derived content are assigned that content by intentional agents who already have thoughts with meaning. Underived content arises from conditions that do not require the independent or prior existence of other content, representations, or intentional agents. So, for example, if minds evolved, the first mind did not acquire its thought content from any other mind (there were no others). The hypothesis that there exist these two types of content raises obvious questions. How exactly does non-derived content emerge? Or, under what conditions does an object 'X' bear some non-derived content X? How is derived content actually derived? Or, under what conditions does an object 'X' bear some derived content X?

Dennett (1990) has this to say regarding the first question:

Where, though, do we get our 'original' and underived intentionality? From God, as Michelangelo suggests? Or is it a mistake to ask such a question? After all, if our intentionality is truly original ('ursprünglich' as a German might say) then it is not derived from anything at all, but rather an ultimate font of meaning, an *Unmeant Meaner*. I take it that this is Searle's view. If it is not, then he certainly owes us an answer to the question of whence cometh our own intentionality. (p. 54)

Dennett's flippant comments here may suggest that he does not have the relevant notion of underived content—roughly, content not derived from antecedently existing content, representations, representational capacities, or intentional capacities. Yet he clearly does, as can be seen by his (disapproving) reference to “unmeant meaners.” In addition, Dennett has written critical responses (1987a,b) to both Dretske's (1981) early work on informational semantics and Fodor's *Psychosemantics* (1987), which contained an early version of an asymmetric causal dependency theory of meaning. More important than taking Dennett to task for some infelicitous comments is the need to draw attention to the literature on naturalized semantics. This literature attempts to explain, not how non-derived content can come from nowhere, but how non-derived content can come from the content-free. This body of work attempts to explain the naturalistic conditions that might constitute a basis for non-derived content (for an excellent introduction, see Stich & Warfield, 1994). So, for example, on the asymmetric causal dependency approach developed by Fodor (1987, 1990), a thought symbol 'X' may non-derivedly come to mean X by the following conditions:

'X' means X, if:

- (1) Xs cause 'X's is a law.
- (2) Some 'X's are actually caused by Xs.
- (3) For all Y other than X, if Ys qua Ys actually cause 'X's, then Ys causing 'X's is asymmetrically dependent on Xs causing 'X's.
- (4) There are some non-X caused 'X's.

One can worry about exactly what Fodor has in mind for his theory and the viability of the theory as a theory of non-derived content, but the ambition to have content emerge from the previously non-contentful is clear (see Adams & Aizawa, 1994).

As another example, on the picture theory of representation, developed by Cummins (1996), non-derived meaning arises in something like the following way:

‘X’ means X if:

- (1) ‘X’ is part of a relational structure R.
- (2) X is part of a relational structure C.
- (3) The relational structure R is isomorphic to the relational structure C.
- (4) ‘X’ is mapped to X under this isomorphism.

As with the Fodorian example, one can worry about exactly how one is to interpret Cummins’s theory and its viability as a theory of non-derived content (for a critique, see Aizawa, 2003), but the ambition to have content emerge from the previously non-contentful is again obvious.

And on Dretske’s theory (1988), the semantic content of thoughts is not derived from prior semantics. It arises due to acquisition of indicator function. ‘X’ comes to mean X when ‘X’s come to have the natural function of indicating Xs. If ‘X’ is a non-semantically derived symbol in the mind, it acquires a non-derived content when it comes to mean X. It means X via a process of neural recruitment for its ability to indicate Xs. Indicating the presence of Xs is not the same as semantically meaning Xs. The natural recruitment process is not driven by a mind or intelligence but by the same kinds of natural causes that drive natural selection for the function of the heart or kidney. So how ‘X’s come to mean Xs on this view is a theory of non-derived content—i.e., the content is not derived from other states with semantic content. To repeat, what matters in the present context is that there are now serious attempts to explain how non-derived content might arise.

The question about derived content, i.e., how the process of content derivation works, has been less thoroughly explored by philosophers of mind (but, cf. Lewis, 1969). Dennett suggests that mere willing can endow something with a derived content (Dennett, 1990, p. 54). Perhaps ‘X’ means X if some normal human being wills that ‘X’ means X. Perhaps a string tied around one’s finger means to remember to bring home coffee, if some normal human being simply wills to have the string bear this meaning. Yet, mere acts of will appear to be insufficient to make sense of all cases of derived content. Perhaps merely willing could work for a string tied around a finger, but it could not make a word meaningful to others. Nor is it clear that some sort of performative act (e.g., saying, “By this flag I will mean peace and nothing else”) is sufficient to establish a corresponding derived meaning. Witness how such performatives might fare in attempts to assign new meanings to the Confederate battle flag or a burning cross.

Suffice it to say that there is no consensus in the philosophical community on what conditions enable ‘X’ to bear non-derived content X or on what conditions enable ‘X’ to bear derived content X. *A fortiori*, there is no consensus on a more

subtle question: If 'X' means X in virtue of satisfying the conditions of some true theory of derived content, does this preclude 'X' from meaning Y (or X) in virtue of satisfying the conditions of some true theory of non-derived content? If we pitch the question quite abstractly, asking whether satisfaction of conditions on derived content precludes satisfaction of conditions on non-derived content, we can see that the answer is not obvious barring a specification of these conditions. Surely it is not an obvious logical or conceptual truth that satisfaction of one set of conditions precludes satisfaction of the other. For all anyone knows, satisfaction of conditions on derived meaning does not preclude the satisfaction of conditions on non-derived meaning.

Dennett is not entirely silent on this matter. He suggests, "No artefact of ours, no product of AI [Artificial Intelligence], could ever have original intentionality simply in virtue of the way it conducted itself in the world or in virtue of its 'formal' design" (1990, p. 57). Here Dennett is echoing Searle's (1980) view of the matter. Dennett's case for this begins with consideration of an encyclopedia with derived intentionality. Even if we were to make an electronic version of this encyclopedia with automated question-answering capacities, thereby giving the system additional "formal" design features and additional capacities for interacting with humans, it would still contain only derived content. Similarly, no matter how we program a chess-playing program, he supposes, it will have only derived intentionality with derived goals of defeating human opponents, concealing what it knows from its opponents, and tricking its opponents. Okay. Suppose, for the sake of argument, that these suggestions are on the right track, that no artifact could have original intentionality merely in virtue of its programming or the way it interacts with other agents. One can concede the force of Dennett's argument and the conclusion he draws from it, while still maintaining that there are conditions other than "formal" properties of machine design and conduct in the world that enable a machine to have original (underived) content. Other conditions, such as those in Fodor's (1987) asymmetric causal dependency theory, or Cummins's (1996) picture theory of representation, or Dretske's (1988) indicator function account might suffice. So, if we review Dennett's paper, we find that he does not really even attempt to establish a conclusion as strong as that artifacts cannot have original intentionality or that 'X' has derived content precludes its having non-derived content. Such a case does not appear to be on his philosophical radar. So, as things stand, it appears that one cannot simply move from the observation that 'X' is an artifact and has derived content to the conclusion that 'X' lacks non-derived content. Such a move requires more argumentation than Dennett has provided.

So, to lay our cards on the table, here is how we think about derived and non-derived content in machines. This is an exposition of our view, not an argument for it. Suppose we wanted to build a genuine thinking machine. To do this, we believe we would have to design the machine in such a way that the machine's symbols mean something to the machine in virtue of satisfying some conditions for non-derived content, conditions such as those in Fodor's or Cummins's or Dretske's theory of meaning. The symbols in the mind of the machine would likely have meanings to

us, insofar as we have to understand the machine and its design to build it. Yet, because the machine is designed to meet the conditions of a true theory of non-derived content, the symbols would also have meanings for the machine. Indeed, somewhat surprisingly perhaps, the symbols in the machine 'X,' 'Y,' and 'Z' could mean X, Y, and Z in virtue of satisfying conditions for both derived and non-derived content. It is in virtue of satisfying the conditions of a theory of non-derived content that the meanings of the machine's symbols would not be semantically derived or dependent upon us. The syntactic items in the machine would be causally produced by us, but they would not have their contents derived exclusively from us.

So much for stage setting. We are now ready for Dennett's principal argument, an argument that begins with an analogy (1990, p. 57). Dennett invites us to imagine someone who creates a robot to carry her in suspended animation for four centuries. To insure the passenger's survival, this robot is designed to navigate about the world seeking energy sources and avoiding danger. It is further designed to interact with humans and perhaps other robots in order to further the survival of its designer. Such a robot, Dennett supposes, has only derived intentionality or derived content, since it is a *mere artifact*. As noted above, the claim that the robot's content is merely derived does not follow as a logical or conceptual truth from the fact that the robot is an artifact, but since it is Dennett's example we will simply grant that the robot does not satisfy the conditions of any true theory of non-derived content. Once we concede that the robot has merely derived content, however, the game is supposed to be over. Since the derivation relation that obtains between the robot and the designer of the robot is supposed to be the same as the semantic derivation relation that obtains between a normal human being and her genes, we are apparently supposed to see that human cognitive content is derived. From this, we are also supposed to get the other conclusion, that humans don't have original (i.e., underived) intentionality.

As we see it, Dennett does not establish either of these conclusions. Let us first consider the idea that all human cognitive content is derived. It turns out that Dennett is somewhat equivocal here. At times he appears to think that human cognitive content derives from our genes in some sense of being a product of natural selection during our evolutionary history.<sup>2</sup> At others times, he appears to think that human cognitive content derives from our genes in the sense of being a causal product of the developmental process.<sup>3</sup> So, a thorough refutation of Dennett's argument will have to address both the phylogenetic and the ontogenetic interpretation. We will try to be thorough.

For Dennett's argument to show that all of our cognitive content is phylogenetically or ontogenetically derived from our genes, the sense in which our cognitive content derives from our genes must be the same as the sense in which the contents of stop signs, flags, and words of English derive from human minds. Although a maximally compelling case for the difference between content derivation and these ontogenetic and phylogenetic derivations might involve having secured a true theory of content derivation, one can still draw attention to significant differences between these types of derivations. Dennett is simply wrong to maintain

that “our intentionality is highly derived, and in just the same way that the intentionality of our robots (and even our books and maps) is derived” (Dennett, 1990, p. 62).

Consider the fact that organisms derive from their genes in a developmental sense. Content derivation is essentially a meaning conferring process, where ontogenetic development is not essentially a meaning conferring process. Ontogenetic development does not necessarily make an organism, or any of its parts, represent anything to, or mean anything for, its genes. A penguin does not mean anything to the DNA that was causally relevant to its ontogenetic development. A penguin does not stand for or mean anything to its genes. A stop sign stands for something and means something to many humans. If you don’t like penguins or their ontogeny, take plants or fungi as your example.

Next consider the idea that humans are derived to some degree by natural selection. Where content derivation is essentially a semantic process, evolution by natural selection is not. Humans and penguins don’t mean anything to natural selection, nor the other way around. Humans and penguins aren’t representations that natural selection uses. If you don’t like human or penguin evolution, again take plants or fungi as your example. Mother Nature doesn’t use humans, penguins, plants, or fungi to represent anything. So there is nothing like the *assigned meaning* model of derived content going on in natural selection.

Still Dennett is clearly suggesting that since natural selection is intentional, and our thought symbols may have derived from processes of natural selection, then our thoughts too have derived content. The idea is that our thoughts have derived content because they are derived from selectional processes that are intentional. To this we say: Causally derived? Yes. But semantically derived? No. There are causal processes involved in natural selection. But there are no minds or semantic contents involved in the *selecting*. Natural selection is not a conscious process. The causal processes involved in natural selection do not have semantic meaning (for example, nothing in them can be falsely tokened). It is true that there can be *selection for* a trait or characteristic. But this sense of *selection for* is not the same as conscious, purposive selection for (e.g., a new car). In the latter case, one has to have the idea of what a car is and a desire for one. In natural selection for, say, a heart, there was no conscious idea of what a heart was or a desire that creatures have hearts in the process of natural selection. The foregoing considerations give us reason to think that the notion of derivation used with regard to derived content is not the same as the notion of derivation found in ontogeny or in natural selection.<sup>4</sup>

Here is another difference between the kind of derivation found in derived content versus the kinds of derivation found in ontogeny and phylogeny. Where content derivation is essentially a process of conferring meaning, ontogenetic and phylogenetic derivations are not meaning conferring. They are, instead, a species of causal production. A bird is developmentally derived from its genes in the sense that the genes are causally relevant to its growth and existence. A bird is historically derived by natural selection in the sense that natural selection is causally relevant to its form and existence. Yet, an item bearing derived content need not be causally

produced by a cognitive agent for whom it is meaningful. It is not the blue-collar manufacturer of a stoplight who makes the red light mean stop and the green light mean go. It is not the person who waves the white flag that makes it mean what it does. It is not the typist who makes the letters on the page mean what they do. Though there are cases where making (causal production) and making meaningful (deriving content) coincide, the process of making an object is distinct from the process of making an object meaningful. The process of tokening a representation is distinct from the process of making that token meaningful.

Even if there is a sense in which DNA encodes instructions (blueprints) from which are built symbols in the brain, this does not mean that those symbols in the brain acquire their semantic content from the DNA. First, the structure in the brain, formed during development and which will become a symbol, may have to acquire its content by causal interaction with information from the world (via one of the theories we referenced above). The genes may endow us with “belief boxes” but it hardly fills those boxes with beliefs that *p* (e.g., beliefs about computers,). Second, even if DNA represents, it cannot be falsely tokened (such as a thought or belief, or intention). So there is nothing in DNA or its causal activity during development that resembles the way that meaning is assigned by a human mind to an artifact or symbol. There is nothing that approaches the notion of derived content with which we (and Dennett) are working. So, there are good reasons to think that the derivation of the human mind from the human genome is unlike the derivation of derived content from prior content.

Suppose, however, that Dennett could solve all these problems, that he could secure the claim that there is human cognitive content that is either an ontogenetic or phylogenetic artifact of the human genome—i.e., suppose that he could secure the claim that there are things in the human brain that have the derived content they do in virtue of satisfying some set of conditions on derived content. Still, from what we have said before, one cannot infer just from the fact that human brain states that have some *derived* contentful states in virtue of one set of conditions, that human brain states lack *underived* content in virtue of another set. Additional argumentation must be given to support the idea that satisfying the conditions on derived content for some objects precludes the satisfaction of any set of conditions on underived content. Yet, Dennett provides no such argumentation.

So far, we have rejected Dennett’s argument for the view that human cognitive content is derived and that human cognitive content is not original. Suppose, however, that we concede these two claims to Dennett. Even with these generous concessions, Dennett still has no arguments for his other claims about original and derived intentionality. Take the claim that there is no distinction between derived and non-derived intentionality. Even if Dennett were correct in his view that human cognitive content is derived from our DNA or from our evolutionary history and that we lack original intentionality, his robot thought experiment does nothing to show that the distinction is false or ill considered. Dennett seems simply to be mistaken about the consequences of his own thought experiment (cf. 1990, p. 56).

Finally, there is Dennett's claim that there is no such thing as original intentionality and, in fact, that the notion of original intentionality is "incoherent" (1990, pp. 54, 62). Suppose that Dennett's robot thought experiment did secure the conclusion that humans have only derived intentionality. Still, that does not show that there is no original intentionality. Wouldn't this just push the originality back on either DNA or Mother Nature? So, why wouldn't DNA or Mother Nature have original intentionality? Dennett *claims* that they don't have original intentionality in the sense we have been considering; instead, they have the intentionality they do in virtue of the intentional stance (1990, pp. 59–60). But, there is no argument for this because Dennett does not explain this derivation. Further, if there is content within the "intentional stance," how did it get there? Nowhere in Dennett's corpus does he come to grips with this question (see Adams, 1991).

So, as far as we can tell, in "The Myth of Original Intentionality" (1990), Dennett has given no good reason to think that humans have only derived content. Nor has he given any reason to think humans lack original (non-derived) content. For that matter, he has not given any reason to think that non-derived content is a myth or to doubt the derived/non-derived distinction.

## Notes

- [1] We think thoughts have non-derived semantic content. We believe natural language, street signs, and other artifacts derive their content from thoughts. We can imagine someone thinking that it goes the other way around—that there is meaningful language first and thoughts derive their semantic content from language. Maybe one thinks that vervet calls have semantic meaning (predator above or predator below) and do not derive this meaning from voluntary, intended mental states, but from involuntary responses. Perhaps one thinks that "language happens" replete with meaning and then becomes internalized and that is how thoughts acquire their content. This would put things the other way around from the way we believe things go. Nonetheless, if semantic content of natural language did not derive from semantic content of thoughts, there would still be non-derived semantic content of just the sort Dennett says cannot exist. It may even be that the causal conditions of one of the naturalized semantic theories we refer to below could be adapted to explain the origin of the semantically non-derived content of language. So while we offer thoughts as the paradigm cases of things with non-derived content, we could adapt our arguments for non-derived content to apply to whatever has it—thoughts, language, or whatever.
- [2] Cf. "We may call our own intentionality real, but we must recognize that it is derived from the intentionality of natural selection" (Dennett, 1990, p. 62).
- [3] Cf. "But this vision of things, while it provides a satisfying answer to the question of whence came our own intentionality, does seem to leave us with an embarrassment, for it derives our own intentionality from entities—genes—whose intentionality is surely a paradigm case of mere *as if* intentionality!" (Dennett, 1990, p. 60).
- [4] Perhaps the human ontogenetic process gives rise to some innate representations and innate knowledge, but they are not representations for the genes. Such knowledge does not consist of representations genes use.

## References

- Adams, F. (1991). Causal contents. In B. McLaughlin (Ed.), *Dretske and his critics* (pp. 131–156). Oxford, England: Basil Blackwell.

- Adams, F., & Aizawa, K. (1994). Fodorian semantics. In S. Stich & T. Warfield (Eds.), *Mental representation: A reader* (pp. 223–242). Cambridge, MA: Blackwell.
- Aizawa, K. (2003). *The systematicity arguments*. Boston, MA: Kluwer Academic Publishers.
- Cummins, R. (1996). *Representations, targets, and attitudes*. Cambridge, MA: MIT Press.
- Dennett, D. C. (1987a). Evolution, error, and intentionality. In *The intentional stance* (pp. 287–321). Cambridge, MA: MIT Press.
- Dennett, D. C. (1987b). Review of the book *Psychosemantics*. *Journal of Philosophy*, 85, 384–389.
- Dennett, D. C. (1990). The myth of original intentionality. In K. A. Mohyeldin Said, W. H. Newton-Smith, R. Viale & K. V. Wilkes (Eds.), *Modelling the mind* (pp. 43–62). Oxford, England: Oxford University Press.
- Dretske, F. (1981). *Knowledge and the flow of Information*. Cambridge, MA: MIT Press.
- Dretske, F. (1988). *Explaining behavior*. Cambridge, MA: MIT Press.
- Fodor, J. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.
- Fodor, J. (1990). *A theory of content and other essays*. Cambridge, MA: MIT Press.
- Lewis, D. (1969). *Convention*. Cambridge, MA: Harvard University Press.
- Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3, 417–458.
- Stich, S. & Warfield, T. (Eds.) (1994). *Mental representation: A reader*. Cambridge, MA: Blackwell.