

Dire Straits for Dennett. How Not to Talk Your Way Past Human Intentionality

*By Nikolaj Nottelmann,
University of Southern Denmark*

Abstract: Daniel Dennett has produced a large, thought-provoking and often highly entertaining body of literature on human intentionality. Even if Dennett's position is not easily characterized or strapped down in traditional terms, one central theme is recurrent: Human intentionality is nothing special. In principle it may be accounted for in terms similar to those in which we defend ascriptions of intentional states to simple animals, or even machines. The present paper disentangles two lines of thought behind this stance on human intentionality: A fictionalist- and a constructivist position not always held clearly apart by Dennett himself. The paper argues that neither line of thought ultimately succeeds in demystifying human intentionality.

At several instances of his long career Daniel Dennett has voiced serious skepticism towards contemporary metaphysics as an academic discipline and its current systems of categorization.¹ Accordingly, he has persistently refused to self-apply any fashionable metaphysical tags, and, on the few occasions where seemingly he has explicitly embraced “instrumentalism” or “behaviourism”, been quick to temper that impression. His overall attitude seems to be that his examples and expositions speak clearer and louder than any label. There is a delightful historical resonance to Dennett's starch impatience with academic metaphysics, echoing the brazen fanfares of the Vienna Circle filtered through Quine's very American temperament. But, what is one to do, if one is a fan of Dennett's writings and is still unwise enough to see something worthwhile in contributing to ontology; “the most Olympian feat in metaphysics”.² In particular, what is one to do with peculiar Dennettian pronouncements like: “We think beliefs are quite real enough to call real just as long as belief-talk measures these complex behaviour-disposing organs as predictively as it does”³? Below I shall discuss Dennett's talk of belief-talk.

¹ See e.g. Dennett (1993, 212) where Dennett declares that „he does not feel obliged to engage in the ontological enterprise”, until “the professional ontologists” agree on a range of basic matters.

² Thus characterized in Dennett (1993, 205), not without a stint of irony.

³ Dennett(1998, 114).

A telling aspect of Dennett's way of doing philosophy is that, as impatient as he is with metaphysics, the more impatient he seems to be with the diverse inventions used by academic philosophers of language in order to mollify metaphysical disputes. In his engagements with metaphysical matters, Dennett resists the aid of substitutional quantification, expressivism, semantic minimalism etc. Thus, for all his resistance to strong forms of realism and objectivism concerning intentional states, Dennett is not a semantic irrealist about intentional talk either: It seems a firm presupposition in his work that attributions of beliefs or desires have a content fit for representing aspects of reality⁴, can stand in genuine logical relations etc. Also, he does not attempt to argue that by a sentence like "Dennett believes that robots will sometime be conscious", I could purport to talk about a certain state of affairs without thereby talking of a certain belief. In Dennett's view folk-psychological talk purports to talk about beliefs and desires as actual wordly occurrences.

But if one thing is clear from Dennett's writings, according to him intentional talk does not succeed in this regard, at least not in the sense that talk of broadly physical objects and occurrences succeed. Dennett's self-proclaimed commitment to "the Natural Ontological Attitude" a.k.a. "conservative standard scientific ontology and epistemology"⁵, clearly makes him view talk of e.g. brain tissue as ontologically innocuous in a manner in which talk of beliefs and desires is not. Attempts by e.g. Robert Brandom to set Dennett's intentional and physical stances on an equal ontological footing in terms of propriety of explanation are clearly misguided.⁶

Thus, given the above observations Dennett needs to account for the status of intentional items, not least how their place in the world differs from that of unproblematic material objects. One major current running through his work on intentionality may properly be labelled "fictionalism". It is the side of Dennett that want to relegate intentional talk to talk in the "as if" mode. This trend is highly visible in his early programmatic paper

⁴ C.f. Wright(2002, 207).

⁵ Dennett(1993, 205)

⁶ See e.g. Brandom(1994, 58). See also e.g. Dennett(1971, 88): "From this [the physical stance] our predictions are based on the actual physical state of the particular object, and are worked out by whatever knowledge we have of the laws of nature." Dennett is strongly unwilling to say anything remotely similar concerning the intentional stance.

Dennett(1971), where it is made clear that, adopting the intentional stance, we merely regard an object *as if* it had beliefs and desires and were rational.⁷

To appreciate the value of Dennett's contribution in this regard it seems salient to distinguish at least four versions of fictionalism⁸ concerning intentional talk:

F1. The view that intentional *talk* has a discursive status equivalent to the telling of fairy tales in our culture: Only the naïve respond to such talk in the expectation that it refers to anything like actual wordly occurrences.

F2. The view that intentional *talk* ought to have a discursive status equivalent to the telling of fairy tales in our culture: Only the naïve may be excused for responding to such talk in the expectation that it refer to anything like actual wordly occurrences.

F3. The view that seeming *matters of fact* concerning intentional occurrences such as someone's believing or desiring something are at bottom mere fictitious matters.

F4. The view that seeming *objects* within the intentional realm such as beliefs and desires are merely fictitious entities.

Here F1 is obviously false and deserves no further attention. However, F2-F4 are independent theses, each demanding some interest in its own right. E.g. as the work of Jennifer Hornsby demonstrates, there is room to hold F4, while denying F2 and F3⁹. I take it to be distinctive of Dennett's position with its unfriendliness to semantic irrealism, that he vehemently denies F2 without taking this in the least to compromise F3 and F4: We are all excused for talking of beliefs and desires as existent, since this mode of talk comes so handy to us. This is most visible when he talks of "the inescapable utility of the intentional stance".¹⁰ Thus, on a salient fictionalist reading (affirming F3 &

⁷ Dennett(1971, 92).

⁸ For purposes of simplification I shall take license to ignore the finer differences between fictionalism and outright eliminativism, which will vary according to one's take on the ontology of fictional objects.

⁹ See in particular Hornsby(1997)

¹⁰ Dennett(1987, 314). He here talks of utility in a biological context, but it seems clear that he holds the intentional stance to be as "inescapably useful" in almost any other context as well, not least in psychology, be it folksy or scientific.

F4), Dennett seems to owe us an explanation why talk of mere fictions as existent can come so natural to us and even be incredibly useful. Read in this vein, certain considerations in Dennett's writings may be taken as a way to answer this question by way of analogy. E.g. when wildly protesting against Jerry Fodor's suggestion that, if beliefs and desires are at most fictional objects, they can play no role in evolutionary theory, Dennett offers one of his favourite examples, the "exactly parallel"¹¹ example of centres of gravity. Centres of gravity like beliefs, Dennett maintains, are "abstract objects", "real" in any interesting sense only in the sense of being predictively useful posits in our scientific endeavours¹². As Dennett is satisfied to observe, still it makes perfect sense, even in a scientific context, to say that a particular boat's low centre of gravity prevents it from capsizing.¹³ The morale we are invited to draw is presumably this: Beliefs are no less metaphysically mysterious than centres of gravity, and obviously appeals to *those* things bring us all sorts of benefits.

On a closer inspection, Dennett's preferred analogy crumbles. As John Haugeland has pointed out, a centre of gravity is not really that mysterious after all. It is a spatial point, a location in the physical universe¹⁴, hence an entity that should be acceptable to even the most austere physicalist. Only it is a spatial point invested for the sake of calculus with the property of having a certain mass, namely the entire mass of the object whose centre of gravity it is. Obviously, unless one is a rambling idealist, neither the point in space nor the mass allocated to it is in the least fictional: A centre of gravity is a perfectly real entity invested with a perfectly real property, only a property, which happens not to be instantiated by itself. In terms of scientific utility, a centre of gravity is rather like the ideal patient of a medical text book, assumed for ease of calculation to react to anaesthesia in a way exactly typical of her age and weight. In the contrasting case of belief we have neither an "innocuous" object, nor an "innocuous" property to attach to it in our fiction or abstraction. Cf. Dennett, we may of course try to make room for persons

¹¹ Dennett(1993, 216).

¹² Dennett(1998, 96-97).

¹³ Dennett(1993, 216).

¹⁴ Haugeland(1993, 55)

within our “Natural Ontological Attitude” and attempt to construe beliefs as fictitious properties of such an entity, but such properties hardly aspire to the degree of “conservative” respectability enjoyed by mass or motion.

The problem inherent in Dennett’s analogy between centres of gravity and beliefs point to a deeper problem in the type of fictionalism ascribed to him above. Cast in this mould Dennett sets off on the wrong foot altogether. For some posit to work as a fiction we arguably need to make sense of the idea what it would be for that object to *actually* exist. It demands only a little imagination to consider what it would take for a single point to exert the pull on surrounding objects characteristic of a certain mass. Likewise we can well consider the nasty implications of fire-breathing fairy tale dragons roaming our lands. But what, supposed that people do not “really” have beliefs, would be the consequences of their actually having them? Would the world change one bit? In other words, we are not much helped by the idea that beliefs are theoretical posits, if we are not told which kind of posit they are. We have no idea what holding them to be fictional entities amount to, for we have no idea about what entities beliefs are in the first place.

Being fair to Dennett, he has persistently chided those philosophers (particularly John Searle) who have insisted on a radical difference between, say, acting as if you believed that a tiger was in front of you, and Really (with a capital r) believing that a tiger was in front of you. All the difference there could be would be degrees of success in explaining your behaviour given a framework positing that belief as one of your propositional attitudes. Even if one may wonder what purpose examples like that of the centre of gravity serves outside of a fictionalist framework, fictionalism thus seems a wrong - or at least a highly uncharitable - interpretation of Dennett’s view on intentionality.

Instead, we may try to explicate Dennett’s view using what seems like the most straightforward reading of the puzzling assertion quoted above that beliefs are “real enough to call real just as long as belief-talk measures these complex behaviour-disposing organs as predictively as it does”.

Leaving aside the bemusement one may feel concerning how something may be “real enough to call real” (is my bicycle real enough for that? Is my

style of riding it?), the statement easily lends itself to a *constructivist* interpretation: Beliefs are perfectly respectable worldly entities (hence not fictions), but they are entities owing their existence to someone adopting the relevant stance towards the relevant believer. Not that the interpreter is causally responsible for the belief's existence like I owe my existence to my parents. Rather, we may say that the belief is (in a non-causal sense) constituted by the interpreter (given further salient circumstances). This, as is clear from the quotation, makes belief-talk *objective* in a certain sense: Not just any belief-ascription will constitute a belief: It needs to ascribe the belief to an entity harbouring relevant "behaviour-disposing" organs, further the ascription needs to be part of a general strategy for predicting the entity's future and the strategy needs to be sufficiently successful. Still, the objectivism inherent in these constraints is a far cry from any "industrial strength Realism"¹⁵, according to which the existence of a belief is entirely independent of anyone ascribing anything to anyone, or at least is only causally dependent on such ascriptions (in the weak sense that many of our beliefs are in fact caused by our participation in ascriptional practices.)

Despite Dennett's explicit unwillingness to offer any "theory of belief" back in 1982¹⁶, it is hard to resist ascribing one to him from 1991. However, the theory outlined above seems vulnerably to several forceful objections, some of which have already been extensively ventilated in the literature. Here I shall concentrate on two objections, the latter of which seems fatal.

The first objection concerns Dennett's appeal to predictive utility as part of the *constituens* for a belief. The trouble is that, arguably, a narrow focus on prediction threatens to deprive beliefs of a salient mark of reality, namely causal powers. In an important footnote to his 1991 paper "Real Patterns", Dennett tries to counter this objection: "If one finds a predictive pattern of the sort just described one has ipso facto discovered a causal power – a difference in the world that makes a subsequent difference testable by standard empirical methods of variable manipulation."¹⁷ It is not altogether

¹⁵ Dennett's term for views like Fodor's, according to which beliefs are sentence-tokenings written in the brain's machine code.

¹⁶ Dennett(1987, 118). Date referring to first printing.

¹⁷ Dennett(1998, 112).

clear what Dennett alludes to by “standard empirical methods of variable manipulation.” However, it seems obvious that putting items in a pattern exploitable for predictive purposes does not in the least presuppose a causal connection between those items. To take a particularly banal example: I may reliably predict the stand of my neighbour’s barometer by reading off my own. I may even test this predictive power under a wide array of circumstances: It holds in fair weather, in stormy weather, at dusk, at dawn etc. etc. However, our barometers are not causally connected. They are mere co-variants, designed to move in unison with the atmospheric pressure. Any *serious* manipulationist theory of causation will thus have to propose “methods of variable manipulation” which effectively screen off the influence of factors responsible for co-variation, such as the atmospheric pressure in our tedious example, without thereby ruining the connections responsible for genuine causal interaction.¹⁸ One way to do this could be manipulating my barometer by exposing it to strictly local changes in air pressure using a pressure chamber and observe the failure of my neighbour’s barometer to follow suit. However, no such experiment could be fully conclusive: One could always (albeit crazily) protest that the pressure chamber also ruins the exchange of a strange sort of radiation between the barometers, the factor responsible for their genuine causal interaction under normal circumstances. This shows that the construction and interpretation of such tests is always informed by basic assumptions concerning what it makes sense to test for, not least an expectancy range concerning the causal powers of the entities tested.

Thus, for all it is worth, on Dennett’s 1991 constructivism a belief may owe its predictive utility entirely to co-variance, rather than to causal efficacy. Dennett’s talk of belief-talk “measuring complex behaviour-disposing organs” only seems to make this a very live option, for co-variance, not significant causal efficacy, is exactly what we want and normally get from our measuring equipment. Further, even given a slightly more sophisticated manipulationist account of causality, it is far from clear, exactly how we could test for causal efficacy. How could we reliably screen off factors responsible for the covariance of intentional states, constructivistically construed, without

¹⁸ This is a major theme in Woodward(2003)

shutting down the intentional system altogether? And what would be the expectancy range on which we could mould our experiments?

The second objection I shall pursue here is hardly more original than the first. It is the oft-voiced worry that a constructivism of the sort envisaged above is threatened by a vicious regress: In order to adopt the salient productive intentional stance towards a suitable system, the interpreter must itself be endowed with intentionality, it must harbour beliefs about the state of the interpreted system, desire to investigate it etc. However, if those states can only exist due to the operation of yet another interpreter, this must stop somewhere, especially since, as often admitted by Dennett, it is only a small minority of intentional systems (primarily the human ones) which are capable of adopting the salient stance. But halting the regress would seem to require the postulation of an entity endowed with intentional states not constituted by any stance, an abominable idea to Dennett's mind.¹⁹

To his credit, Dennett has repeatedly tried to counter this objection. Most of his counter-offensive takes the form of an attack on the idea of "original intentionality", a mystical power reputedly imagined by John Searle and others, in virtue of which the human brain is able to cause intrinsically contentful states, owing neither their status as intentional nor their particular contents to the activity of any interpreters. I am entirely sympathetic to Dennett's idea that brain tissue does not have a principled monopoly on an entire category of causal powers, namely the power to produce "real" contents. However, I also sense that Dennett's counterattack misses the real bite of Searle's criticism. What I sense is that Dennett fails to observe a crucial distinction: The distinction between content-conferrers and content-interpreters. Or to put it differently: Between content-fixers and content-constituters.

A closer look on Dennett's most elaborate counter-attack should make this clearer. In his 1987 "Evolution, Error, And Intentionality", Dennett sets out to set his favourite Prügelknaben - Searle, Fodor and Dretske - right on these issues. What Dennett hangs his hat on is a lowly soda-vending machine, the two-bitser. He convincingly brings out that the content (if any) of this

¹⁹ Brandom(1994, 60) is only one of many texts voicing this concern.

machine's internal state after a quarter has fallen through its slot, depends not only on the machine's original design, but also on its history of use. E.g. no matter its original design idea, its slot may be a balboa-detector if the machine is successfully used in Panama with the purpose of letting out a soda when a suitable amount of balboas is thrown into it. Now, Dennett argues, any state in nature which is a candidate for interpretation as contentful derives its content from the design and use of the relevant system. In the end, even contentful states of human beings derive their content from our species' history of evolution as well as the individual's (and other individuals') use of her cognitive make-up as produced by that history. From the scientific vantage point of evolutionary theory, attributing design ideas to Nature is inescapable. Thus all intentionality is derived in the sense that all contentfulness derives from the ultimate purposes of nature, and must be investigated by scientific methods. Down the drain goes privileged access to content of own states as well as all the other ghosts of internalism.

We may seriously doubt that the two-bitser's internal states present any fair analogue to the states of human brains responsible for "complex behaviour-dispositions." At least the two-bitser's internal state lends itself to a broadly physical description hardly satisfactory in the human case. How, e.g. account for the fact that a human could have more than half of his brain shot away and still retain most of his memory? However, with a bit of charity it seems clear that Dennett's story, read as a tale of the general constitution of content, escapes vicious regress worries: If all that it takes to fix content is use and design in the grand but minimal sense that Nature is capable of, it seems clear that we all design and use each other all the time. History is one happy holistic content-fixing mess. Still, Searle & Co. seem entitled to protest: Given that you really are a constructivist in the sense outlined earlier above, you have nicely shown where an interpreter should look for evidence when ascribing particular contents to a system. We could agree (at least for this purpose) that she must look not only to the individual's current activity and brain but to her wider history and the history of her species. However, this only points us to the *fixers* of her mental content, once the vital presumption is made that there is any content to look for in the first place. But, given your constructivism

there could only be a content to look for, once someone has adopted the intentional stance, i.e. engaged in the activity of looking for content fixers. Thus, even if the content is *fixed* in a holistic, yet non-vicious way, it is not *constituted* thus. Design and history of use fixes the content only *as seen from* the stance constituting it. And the activity of adopting this constituting stance must stop somewhere. Surely we need not presuppose extraordinary *content-fixing* powers to the human brain in order to get past this worry. We need only attribute a seemingly commonsensical ability: The power to interpret without being itself interpreted. This is a power, which only *humans* and perhaps a few animal species seem to possess. Hard to say if robots will develop similar powers in the future. Even harder to say if we should ever credit them with it.

If this diagnosis is correct, Dennett's constructivism is seriously at odds with his persistent insistence that human intentionality is nothing special. Moreover, his attempts to smear his critics using the ink of a few uncaredful Searlian formulations, has not floored the real worries behind their attacks. Accordingly Dennett the fictionalist and Dennett the constructivist both seem faced with insurmountable opposition. I confess to being at a loss as to fixing a more resistant label on Dennett's position. Perhaps we should give up the labelling game altogether.

However, I shall close this short essay by offering a reason why it may not be so disrespectful after all, even when sticking to a strict third-personal perspective, to see human intentionality as something special, not resting happy (as some of Dennett's critics may perhaps do) with postulating its special status for theoretical purposes. Dennett, in the tradition of Donald Davidson and Wilfrid Sellars, have always insisted on the strong conceptual link between propositional attitude ascription and rationality: Beliefs and desires may only be attributed as (and hence, in any respectable sense, are) the system's *reasons* for actions and inferences. The notion of a system's continually acting contrary to its reasons makes as little sense as a system having reasons not acted upon under any circumstances. A system's reasons may be determinable in several ways: We may try to discern patterns in its behaviour (repeated oogling of the cookie jar reveals a desire for cookies) or we may try to learn of its overall design and history of use. The last strategy

makes the most sense for simple contraptions like thermostats and two-biters. When applied to humans, however, most of the time this strategy appears not only futile, but seriously counterproductive. Surely Nature has designed us, such that most of us harbour a rather stereotype matrix of basic desires and cravings. Most obviously we desire nourishment when not fed for some time, sleep when having been held awake for a day or two, respect from our fellow humans, sex when spending sufficient time with an agreeable partner etc. etc. All of these cravings may be seen to derive from even more basic desires, such as reproduction and survival. Sometimes sales executives and urban planners may base important etiological predictions on ascriptions of those cravings. However, most of the time most of us are interested in much more fine-grained predictions. I am not content to know that my kids will crave food when returning from kindergarten, I want to predict whether they will crave for meatballs or lasagne. Evolution doesn't offer much help here: It may be that the panda's taste for bamboo is fixed by its Ancestral Adaptive Habitat being overgrown by the stuff, but neither meatballs nor lasagnes were around in the stone-age. Moreover these dishes may taste practically the same, yet elicit hysterically different reactions from my kids. That such idiosyncratic desires are not merely hooked up on a fine-graining introduced by language, any parent with babies will testify.

However, human intentionality is not only characterised by its incredible complexity and plasticity. It is also characterised by the fact that it seems capable of processing, questioning, counter-acting or even discarding even its most "innate" and basic reasons: Humans are able not only to form second-order attitudes towards their desires to procreate, eat, drink and sleep; they are able to act contrary to such cravings while still recognizing them within themselves. In principle, no attitude is out of bounds for rationalization, evaluation and questioning. In short the human space of reasons is unbounded, even if, for purposes of predictions and perspicuity it may often be productive to view it as based on a few guiding attitudes. No other type of intentional system encountered in the lab or the field as of yet seems to possess similar features. Even Dennett's famed and sophisticated "survival

robot”²⁰ with all its social and adaptive skills is characterized by the fact that all of its behaviour derives from a single goal or desire: The desire of allowing its dormant human resident to rest unharmed for 400 years. Remove or shake that goal and we shall say that the robot malfunctions. In contrast, we standardly allow that even if most of time it seems prudent to treat severely ortorectic dieters with “corrective” medications, perfectly rational or well-functioning people may go on hungerstrike or eat abominable food out of strong and respectable convictions. In counteracting their basic cravings by inhabiting a contrary rationalization, it is in no way clear that they perform otherwise than they were “designed” to perform.

Thus, in conclusion, unless Dennett shows us that, contrary to appearances, even such aberrant behaviour is informed by basic (unconscious?) goals invested in us by Mother Nature, we shall have a positive reason to regard human intentionality as a special case until a robot shows up exhibiting a behaviour which we cannot help interpreting as indicative of an unbounded space of reasons, or we at least told in convincing outline how to construct such a robot. Above we saw that a fictionalist reading of Dennett’s view of intentionality would not cut any ice. Cast in a constructivist mould we may now credit him with giving a fair construal of the intentionality of systems that lend themselves to the intentional stance, but do not have the power to interpret without being interpreted. On systems exhibiting this power, Dennett has thrown no real light, although he has managed to chase to death quite a few red herrings, and even made great and enlightening entertainment out of it. For this we should be forever grateful.

References

- Brandom, Robert B, 1994, *Making it Explicit. Reasoning, Representing and Discursive Commitment*, Harvard Univ. Press, Cambridge.
- Dennett, Daniel C., 1971, “Intentional Systems”, *Journ.Phil*, Vol. 68, No. 4, 87-106.
- Dennett, Daniel C., 1987, *The Intentional Stance*, MIT Press, Cambridge.

²⁰ Dennett(1990, 57)

- Dennett, 1990, "The Myth of Original Intentionality", in K. A. Mohyeldin Said, W. H. Newton-Smith, R. Viale and K. V. Wilkes, eds, *Modelling the Mind*, Oxford: Clarendon Press 1990, 43-62.
- Dennett, Daniel C. 1993, "Back from the Drawing Board", in Dahlbohm, Bo (ed): *Dennett and his Critics. Demystifying Mind*, Blackwell, Oxford, 203-235.
- Dennett, Daniel C., 1998, *Brainchildren. Essays on Designing Minds*, MIT Press, Cambridge.
- Haugeland, John, 1993, "Patterns and Being", in Dahlbohm, Bo (ed): *Dennett and his Critics. Demystifying Mind*, Blackwell, Oxford, 53-69.
- Hornsby, Jennifer, 1997, *Simple Mindedness. In Defense of Naïve Naturalism in the Philosophy of Mind*, Harvard Univ. Press, Cambridge.
- Woodward, James: *Making Things Happen: A Theory of Causal Explanation*, Oxford Univ. Press, Oxford 2003
- Wright, Crispin, "What Could Antirealism concerning Ordinary Psychology Possibly Be?", *Phil.Review*, Vol. 111, No. 2, 205-233.