Chapter 3
What Was I Thinking? Dennett's *Content and Consciousness* and the Reality of Propositional Attitudes

Felipe De Brigard

> *Now once again is the view I am defending here a sort of*
> *instrumentalism or a sort of realism?*
> *I think that the view itself is clearer than either of the labels, so*
> *I will leave that question*
> *to anyone who still finds illumination in them.*
>
> (Dennett 1991)

Abstract Back in the 1980s and 1990s there was a lively debate in the philosophy of mind between realists and anti-realists about propositional attitudes. However, as I argue in this paper, both sides of this debate agreed on a basic assumption: that the truth (or falsehood) of our ascription of propositional attitudes has direct ontological implications four our theories about their nature. In the current paper I argue that such an assumption is false, and that Dennett had hinted at its falsehood in the first part of *Content and Consciousness*. In an exercise of "counterfactual exegesis", I suggest that, had this point been acknowledged then, this longstanding debate – which still survives to this date – could have probably been avoided.

Back in the 1980s and early 1990s, there was a lively debate in the philosophy of mind between realists and anti-realists about propositional attitudes. On the one hand, there was *intentional realism*, a view primarily defended by Jerry Fodor, who thought propositional attitudes were computational relations between a subject and a real, sentence-like representation in the language of thought. On the other hand, there were a handful of antirealist approaches, with Paul Churchland defending its most radical and influential version: *eliminative materialism*. For most empirically oriented philosophers of mind, this dispute is now obsolete, not so much because it has been settled, but rather because the field has evolved in such a way that many of the terms of the debate are no longer understood as they were back then. For

F. De Brigard (✉)
Center for Cognitive Neuroscience, Department of Philosophy,
Duke University, Durham, NC, USA
e-mail: felipe.debrigard@duke.edu

C. Muñoz-Suárez, F. De Brigard (eds.), *Content and Consciousness Revisited*,
Studies in Brain and Mind 7, DOI 10.1007/978-3-319-17374-0_3

instance, mental representations are now rarely considered sentences in mentalese, and the few contemporary advocates of the language of thought support their views using cognitive and computational neuroscience, rather than using folk psychology as Fodor did (Gallistel and King 2009; Schneider 2011). Similarly, most views on computationalism have matured, and many no longer require the kinds of representational commitments Fodor once demanded (Piccinini 2008). However, in less empirically informed circles, this lack of denouement is taken to imply that the debate has simply remained dormant, and that the arguments deployed in the past are as strong now as they were before (see, for instance, Matthews 2010).

If only for that reason, my current attempt to revive a decades-old debate may not be completely futile. Yet, there is another reason why I think it is worth revisiting this dispute. I have long suspected that both intentional realists and eliminative materialists have based their arguments in a controversial thesis, viz. that the truth (or falsehood) of our ascriptions of propositional attitudes has direct ontological implications for our theories about their nature. This thesis, I believe, was underwritten by a particular take on scientific realism that committed both parties to accept two related assumptions: (1) that truth is as a matter of correspondence between words and things in the world, and (2) that the things named by true theories must exist. This sort of scientific realist stance was not ungrounded, of course. It was motivated by considerations regarding the success and failure of folk psychology. On the one hand, intentional realists took the *success* of our folk psychology as good evidence for the theory's truth, and then went on to suggest that our best theory of the mind should take the syntactic objects of our propositional attitudes as real entities – specifically, mental representations realized in the brain. On the other hand, eliminative materialists like Churchland took the relative *failure* of folk psychology as sufficient evidence for its falsehood, and then went on to suggest that folk psychology was false because it wrongly assumed the existence of unreal entities like beliefs, desires and so forth. The upshot of eliminative materialism was that, being a false theory, folk psychology was doomed to extinction, just like other obsolete theories we used to have.

As mentioned, this dichotomy largely framed the debate about the nature of propositional attitudes in the 1980s and 1990s (Fodor 1985). My contention now is that this was a false dichotomy, and that the debate was ill-construed. Moreover, I believe Daniel Dennett offered an important insight in the first part of *Content and Consciousness* (Dennett 1969; henceforth *C&C*) that, had it been developed, it would have severely weakened the aforementioned controversial thesis. Perhaps because Dennett did not develop this insight in the 1970s, and barely touched upon it when he further articulated his views on the nature of propositional attitudes (e.g., Dennett 1978, 1987, 1991), this important insight went unnoticed. As such, the current essay could be seen as an exercise in "counterfactual exegesis", as I try to develop this Dennettian insight in my own terms, writing on a line of argument that could have been explored years ago, and that might have prevented the development of a debate that, for many, it is now passé. Still, I hope that by incorporating some recent developments in related areas of philosophical research, those philosophers for whom the debate about the reality of propositional attitudes is merely dormant can find new reasons to question its legitimacy.

To that end, I offer an argument in which both eliminative materialists and intentional realists about propositional attitudes turn out to be partially wrong. Briefly stated, the idea is that these views represent two cardinally opposed ways of deriving ontological implications from the same underlying scientific realist assumption, which – I suggest – we would be better off rejecting. In order to make my case, I begin by explaining the origins of the dispute between intentional realists and eliminative materialists. I claim that it spawns from disagreements about a single argument – an argument I dub (inspired by Kitcher 2001) the *success-to-truth argument*. In Sect. 3.2, I talk about eliminative materialism. I argue that Churchland's arguments that folk psychology is false are unsound. I claim then that since there is no good reason to believe that folk psychology is false, the thesis of eliminative materialism cannot really get off the ground. In Part 3, I move on to a critical discussion about intentional realism. My criticism here is two-fold. On the one hand, on the basis of recent developments in linguistics and philosophy, I argue that we do not have enough a priori reasons to believe in the reality of 'that'-clauses' referents. On the other hand, I suggest that Fodor's inference to the best explanation vis-à-vis the reality of language-like mental representations can be challenged as well, casting more doubts on its ontological implications. Finally, in Sect. 3.4, I show how Dennett's insight in *C&C* can be read as anticipating these points, and as offering an alternative strategy to interpret *the success-to-truth argument*, in a way that might relieve the philosopher of mind from awkward ontological commitments regarding the nature of propositional attitudes.

3.1 The Success to Truth Argument

This is the formulation of what I call the success-to-truth argument (STA):

(Assumption) Folk psychology is a theory
(P1) Folk psychology is a successful theory
(P2) If a theory is successful, then it is true. Therefore,
(C1) Folk psychology is true.

Each statement needs some explaining. The Assumption holds that the so-called *'theory'-theory* is true. Barring some idiosyncratic differences in its formulation, the 'theory'-theory can be seen as the conjunction of two claims – the first of which, it appears, is contained by the second (Lycan 2004). The first claim is that mental terms are explanatory; they were inserted into our language to help us predict and explain other people's behaviors. The second claim is that these mental terms perform their explanatory and predictive role in virtue of being part of a theory, a *folk* theory, commonly known as *folk psychology*.

Folk psychology can be first approached by way of an analogy. Folk psychology is to scientific (organized, systematic) psychology as folk physics is to scientific (organized, systematic) physics. As we grow up and learn to navigate the world, we begin to develop an understanding of the structure of everyday objects, about the way in which they behave, how they react with each other or under different conditions, and so forth. In general, folk physics works pretty well. Parental teachings instruct us to estimate with accuracy the trajectory of a baseball, and to then catch or flee accordingly. Less friendly classrooms have taught us to pick out tree branches apt to resist the stress produced by the gravitational force acting upon our well-fed 7-year-old bodies. Thanks to experience, we accrue piles of physical folklore that help us in the business of explaining and predicting the behavior of good old middle-sized objects. Mutatis mutandis, when it comes to folk psychology. Repeated encounters with energetically voiced instructions teach us when it may be wise to cut it out and do as our mother wishes. And our occasional interactions with persons whose behaviors we deemed questionable rightly suggest that they follow some beliefs we do not share. Just as we live in a world packed with middle-sized objects, we also live in a world populated with people. Folk psychology is the understanding we develop to make sense of people's complex behaviors.

It is customary to trace the historical origins of folk psychology back to Sellars' celebrated myth of Jones (Sellars 1956/1963). Details aside, Sellars' fable conveys the idea that mental terms are theoretical terms inserted in our folk psychology to refer to inner, unobservable episodes of others' mental lives – episodes which, are *alleged* to be causally responsible for their overt and observable behavior. Whereas our Rylean ancestors' theoretical repertoire was limited to mere observational/dispositional expressions, Sellars tells us that "Jones develops a *theory* according to which overt utterances are but the culmination of a process which begins with certain inner episodes" (Sellars 1956/1963: 186). These unobservable 'inner episodes' are to be taken as the referents of the theoretical mental terms Jones uses to explain the rich mental life unreachable by the behaviorist. To sum up: the Assumption says that folk psychology is a theory; that just like any other scientific theory, it works in part by introducing theoretical terms; that our mental terms are those theoretical terms; and that, hypothetically, mental terms refer to inner episodes.

The first premise (P1) insists that folk psychology is a successful theory. This premise, in fact, is the Rubicon dividing eliminative materialists and intentional realists. On the one hand, intentional realists suspect that, for the most part, folk psychology works fine. In general, predictions and explanations couched in mental terms seem to work, their generalizations seem to apply to novel cases, and their exceptions seem to be somewhat easily explained away, either by the theory itself, or by pointing at some violation of a ceteris paribus clause. On the other hand, eliminative materialists take folk psychology to be a complete failure, a stagnant science at most, with all sorts of predictive and explanatory shortcomings. Arguments in favor and against (P1) are, therefore, the main topic of the next section.

Finally, the second premise (P2) corresponds to what Kitcher (2001: 177) calls "the success to truth inference". The motivation behind (P2) is the belief that if scientific success is systematic, nothing miraculous must be going on; scientific accomplishments must not to be cashed out in terms of repeated coincidences but – at least intuitively – in terms of truth. Many scientific realists take (P2) as an argument in favor of scientific realism as, allegedly, it is the only view that does not make the success of science look like a sheer collection of systematic miracles. But if this was the only option, one would seem to face an unfortunate dilemma: either one must embrace scientific realism, or one must accept the preposterous thesis that the success of science is pure luck (Votsis 2004). I hope to show, in Sect. 3.4, that some ideas in *C&C* can be read as offering an alternative view upon which to build a rejection of (P2) and a solution to the realism/anti-realism debate about propositional attitudes.

3.2 The Persistence of Folk Psychology

Eliminative materialism, according to Churchland, "is the thesis that our commonsense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience" (1981: 67). The force of this view, I contend, stems from the rejection of (P1). Notice, however, that Churchland needs (P2) to be stronger than the version I provided. He needs the implication in (P2) to be a bi-conditional. As it stands, it may very well be possible for folk psychology to be an unsuccessful theory and yet still be true. After all, there are instances in which certain theories, accepted as true by the relevant scientific community, have failed to produce successful predictions.[1] So Churchland needs (P2) to read:

(P2*) A theory is successful if and only if it is true

This way, if he can prove that folk psychology is actually an unsuccessful theory, its falsehood will be warranted – that is, C1 would be false. To that effect he cites "three major empirical failings of folk psychology" (Churchland and Churchland 1998: 8 [but see also Churchland 1981, 1988]):

(a) Folk psychology cannot explain a considerable variety of psychological phenomena, including mental illness, dreams, and concept acquisition by prelinguistic children, amongst many others.

[1] Here's a possible example of a theory that hasn't produced successful predictions, not because of the falsity of its premises, but because scientists don't know yet how to apply it in experimental or practical situations. Consider Schrödinger's equation. Although it is sufficiently clear which mathematical outcomes could be expected from calculations involving it, some empirical interpretations of such calculations are either unclear or impracticable. Cramer (1988), for instance, suggested an interpretation of the nature of wave equations, such as Schrödinger's, according to which a mixture of real and imaginary numbers is required. The problem is that these complex variables – as the mixed numbers are often called – are written as ± numbers, by virtue of which there are always two possible solutions. Alas, when used in equations involving the behavior of a system in time, the change in sign is supposed to be understood as "reversing" the direction of time, and that – as far as I understand – is still not quite easily interpretable in terms of empirical success. This impossibility, however, purports no harm to the acceptance of the equation as being true, and I suspect there may be similar examples in other areas of physics, perhaps even beyond quantum mechanics.

(b) Folk psychology has remained unaltered for the past 2,500 years, showing no signs of development and many of stagnation.
(c) Folk psychology does seem difficult to integrate with the other disciplines in its theoretical vicinity, like physics, chemistry, biology, and physiology.

The upshot, then, is that folk psychology is unsuccessful and should be deemed as false.

Despite the appeal of these alleged empirical reasons, I think they can be contested. Let us begin with (a). The *main* moral we were supposed to draw from Sellars' myth of Jones was that mental terms were introduced in our folk psychology in order to help us explain the observable complex behavior of other people. More specifically, mental terms were supposed to contribute to the systematization of laws, the purpose of which was to explain and predict the observable behavior of other persons. Now, Churchland considers that folk psychological explanations fail on two grounds: (1) because their theoretical terms depict a "radically inadequate account of our internal activities" (Churchland 1981: 570), and (2) because they prove ineffective when applied to a subset of psychological phenomena (e.g. mental illness, sleep, etc.). However, rejecting folk psychology on the grounds of (1) does not seem fair once we realize that "our internal activities" was not its proprietary domain of evidence and explanation in the first place. When it comes to scientific explanations, it is always important to keep the notion of success relative to the kind of object over which its predictions and explanations are supposed to operate. And it seems clear that in the case of folk psychology these objects are persons. Mental states were never introduced into our folk psychological language in order to stand in place of neural events. It is true that Jones *hypothesized* that theoretical mental terms – perhaps because they *seem to be* referential terms – were supposed to refer to inner linguistic episodes. However, this consideration, as well as any other further considerations regarding the *nature* of such episodes, is going to be either gratuitous or dependent upon subsidiary hypotheses (e.g. that our inner mental life mirrors our overt linguistic life; that mental states are to be correlated with brain states; that there are not non-linguistic inner episodes causally responsible for overt utterances, etc.). If you want to claim that inner episodes are brain events you may provide these subsidiary hypotheses. Nonetheless, for the purpose of the effectiveness of the myth, you need not. For all Jones knows, dualism could be true, the extended cognition hypothesis could be true, in fact, people could even be zombies, and yet folk psychology would still be vindicated. Why? Because the assumption of mental terms – that is, of theoretical terms – serves *primarily* the purpose of systematization: "it provides connections among observables in the form of laws containing theoretical terms" (Hempel 1958/1965: 186). Theoretical terms in our laws are, as it were, operational shortcuts posited in place of a bunch of observational data, which are further used to infer observational conclusions there-from. They do not serve primarily a referential purpose. Therefore, as long as they serve *their* purpose within the laws, whether they fail to refer to our internal neural activities doesn't really matter.

By the same token, to reject folk psychology on the grounds of (2) does not seem reasonable either. Suppose we agree that we have always used mental terms to make sense of people's behaviors. Now, insofar as we have used mental terms in *this* way, psychological explanations and predictions are actually quite successful.[2] In general we are good at interpreting someone else's needs and hopes, what to expect from them given what we know, or even what we don't know. Indeed, the success of folk psychology in everyday life is so ubiquitous that it is "practically invisible" (Fodor 1985: 3). It is true that, at times, our explanations at the folk psychological level seem to fail. But there are failures and there are *failures*. Suppose I ask you to meet me tomorrow at school at 3:00 pm. Suppose further that you say, 'Yes, I'll be there'. From that piece of information I infer that you have formed the desire to meet me at school tomorrow and that you have formed the belief that I will be there at 3:00 pm. Then I put belief and desire together and I predict the following action: that you will go to school tomorrow at 3:00 pm for our meeting. The prediction fails, alas: you forgot the date. What went wrong? Here one has (at least) two options: one can either blame the entire predictive apparatus (i.e. folk psychology), or one can simply argue that your obliviousness constitutes a violation to a tacit ceteris paribus clause. Blaming the entire apparatus of folk psychology on the basis of just one failure seems a bit exaggerated. For one, I can provide an explanation of the failure in terms of the very same theory: if you hadn't *forgotten* the date, my prediction would have worked just fine. Secondly, it is true that similar extrapolations have proved successful in the past (last Wednesday – remember? – you did actually make it to our appointment). Finally, I can also be confident that the new prediction I make right after I talk to you – and you apologized, swore this time you'd be there on time, etc. – is actually going to work, ceteris paribus of course. Then again, maybe the problem is that you may not like ceteris paribus clauses at all. Fair enough. However, if that is so, your concerns can be generalized across the board, for they may actually affect most of our scientific theories (including neuroscience!), not only folk psychology (see, for instance, Lange 2002).

Surely Churchland does not have *those* cases of failure in mind when he claims that folk psychology cannot accommodate certain phenomena. He has in mind *big* failures, like the case of epilepsy. But was this really a failure of folk psychology? It seems to me that epilepsy is merely an exceptional disturbance whose behavioral characteristics are "less psychological" than the prototypical folk psychological phenomena. It is not that epilepsy was not easily explainable by reference to folk psychology's ceteris paribus clauses; it is rather that it was a very odd behavior, like hiccups or somnambulism, and it just did not seem to be the product of typical psychological states. Perhaps that was *precisely* the reason why people introduced demonic possessions to explain epilepsy: since it was not part of the domain of characteristic behaviors folk psychology usually explained, a different discipline was required to do the job. It is true that theology failed to explain the phenomena and that now neuroscience can explain epilepsy all right. However, it is not clear to me how this achievement of neuroscience is supposed to harm the success of a folk theory for which epilepsy was not clearly a proprietary explanandum. For not being able to explain epilepsy in terms of demonic possessions, it is not psychology that should not be blamed, but theology!

[2] Dennett articulated this point, before Churchland's paper, in pieces like *True Believers: The Intentional Strategy and Why It Works* (Reprinted in Dennett 1987).

Similar points can be made regarding other cases of *big* failures Churchland mentions. Take dreams for instance. Dreams do not elicit typical overt behaviors. People rarely behave when they are dreaming. And when they do, their behavior is rarely elicited by any inner episode they are aware of – or, at least, that they could causally respond to in virtue of their content. In that regard, dreams do not seem to be proprietary explananda of folk psychology. Therefore, insofar as they do not belong to the domain upon which folk psychological explanations were supposed to operate, it is unfounded to use them as counterexamples. A similar conclusion can be found in Horgan and Woodward (1985: 402) for whom "There is no good reason, a priori, to expect that a theory like [Folk Psychology], designed primarily to explain common human actions in terms of beliefs, desires, and the like, should also account for phenomena having to do with visual perception, sleep, or complicated muscular coordination" (Horgan and Woodward 1985: 402).

What about (b)? There is a longstanding line of argumentation against the stagnation objection trying to show that, in reality, folk psychology has actually progressed in the past 2,000 years. To that effect, philosophers and psychologists have shown that psychology, at the social and personal levels, makes constant use of belief/desire talk in the process of pushing forward their research programs: "for instance, temperament seems to be more useful in predicting behavior than other sorts of personality traits, according to social psychology; short-term memory holds about seven 'chunks' of information, whether these are numbers or names or grocery items, according to cognitive psychology; and so on" (Schroeder 2006: 69). I think this line of argument is basically right; I'd just add one more point: folk psychology not only proves necessary to the process of concocting research programs but, *more importantly*, to the process of carrying out those programs. It seems undeniable that true ascriptions of mental states are necessary when interpreting and producing neuroscientific data in situ, both inside and outside of the lab. Neuroscientists ought to believe that their subjects' introspective reports are veridical no less than they should trust the word of their co-workers. These intersubjective data would be useless unless we had the network of folk psychology up and running.

Still, there is another reason to be skeptical about the force of (b). 'Development' is a tricky word. In what sense does a theory develop? If developing counts as fostering research programs, then – as Horgan and Woodward (1985) argued – folk psychology has clearly developed. On the other hand, if development means something like "refinement" of a theory's axioms and principles, then I agree: folk psychology hasn't shown that much of it. But then again this sort of "immobility" need not be a sign of failure. It may be a sign of proper functioning instead. If a theory constantly proves unsuccessful and does not undergo revisions and changes, it is right to accuse it of being a bad theory. But if a theory works just fine when it has to, why would we want it to change at all? Consider basic arithmetic. Nobody would reject basic arithmetic on the grounds that it has not undergone any significant changes in the last 2,000 years. Basic arithmetic – the primary school arithmetic that most people operate with – hasn't changed because it works just fine for most everyday tasks. A similar point can be made about folk physics. People keep making the same rough generalizations and predictions about middle-sized mundane objects on the feeble basis of previous successful experiences; yet, so far as quotidian life goes, folk physics works alright and hasn't shown signs of severe alterations. The same goes, mutatis mutandis, for folk psychology.[3]

Let me conclude with a comment about (c). To being with, it seems unclear what the objection amounts to. For the objection to be *really* an objection against the success of folk psychology the following claim should be true: that if a theory A is not integrable to a theory (or a set of theories) B, then A is unsuccessful. Call this claim *the integrability condition*. But what is meant by "integration"? In his 1981 paper, Churchland equates "integration" with the idea that some natural sciences tend toward a "theoretical synthesis" with the physical sciences in which the categories of the former are successfully reduced to those of the latter. But, he says, "F[olk] P[sychology] is no part of this growing synthesis. Its intentional categories stand magnificently alone, without visible prospect of *reduction* to that larger corpus" (Churchland 1981: 75). And it is fair to assume that by "reduction" he means what he meant 2 years before, in his 1979 book: that a theory A is successfully reduced to a theory B so long as two conditions are met: (1) that we can provide a set of rules (so-called "bridge laws") according to which the terms in A are mapped onto terms of a subset of sentences in B, and (2) that the expressions in B which the terms of A were mapped onto are axioms of A (Churchland 1979: 81ff). That way, A will be "contained" in B, i.e. B will explain as much as A explains and more. However, several arguments in the philosophy of science should have convinced us by now that (1) is not the case for most – if not for all – (special) sciences, and that since (2) presupposes the success of (1), (2) may prove impractical as well.[4] Therefore, given

[3] A different concern is to accuse folk physics of being unable to solve puzzles in the domain of scientific (organized, systematic) physics. This is also an unfair claim. Scientific physics deals with highly idealized objects and situations whereas folk physics has a more mundane domain and a very different purpose. I think it would be a mistake to reject folk physics on the basis that its generalizations don't coincide with the generalizations of scientific (organized, systematic) physics. The same, I think, goes for folk psychology. As Andy Clark so eloquently put it once: "Folk psychology may not be playing the same game as scientific psychology, despite its deliberately provocative and misleading label" (1989).

[4] I have in mind the arguments in Fodor's "Special sciences" (1974). For instance, the latter, very briefly, goes like this: a successful reduction of the psychological law like

(1) $S_1x \rightarrow S_2x$

is achieved as long as we can provide bridge laws of the form

(2a) S_1x *iff* P_1x and
(2b) S_2x *iff* P_2x,

guaranteeing the reduction of the psychological predicates S1 and S2 to neurophysiologic predicates P1 and P2 in a law of the form

(3) $P_1x \rightarrow P_2x$.

Alas, this sort of reduction is impracticable because bridge laws connecting type-psychological predicates with type-neurophysiologic predicates are, if not impossible, highly improbable ("an accident on a cosmic scale"). At most, all we can get are correlations between type-psychological predicates with heterogeneous disjunctions of type-neurophysiologic predicates like the correct rendering of *the integrability condition* (if a theory A isn't *reducible* to another theory B, then A is unsuccessful), and given the arguments against the tenability of such reductions, the acceptance of *the integrability condition* required for the success of (c) would force us to reject any theory that proves irreducible as unsuccessful. Sadly, that would include basically all special sciences (not only psychology, but also economics, sociology, and so forth) and some lower-level sciences, like ecology, biology and perhaps neurology. To argue that none of these sciences is successful is preposterous. Irreducibility just cannot be the mark of scientific failure.

Some may object at this point that I am being unfair, as Churchland soon realized that his "classical account of intertheoretic reduction appeared to be importantly mistaken", and offered some "necessary reparations" (Churchland 1985/1992; Churchland and Hooker 1985). Fair enough. I'm willing to assume, for the argument's sake, that his new account actually circumvents the difficulties mentioned above. Still, there is another reason to be suspicious of the idea that reducibility speaks in favor of the success of a theory. If the success of a science is to be accounted for in terms of its explanatory and predictive achievements, then a successful reduction should have a negative effect on the explanatory power of the reduced science. In other words, a reduced science can't provide a better answer for a certain question than its reducing science. But this is hardly the case with folk psychology. Often times, the kind of explanations users of folk psychology require are not neurological. Sometimes we demand historical explanations, or accounts in terms of the environment in which the subject is embedded, or even contrastive answers, as when we wonder why a person decided to do X as opposed to Y. Reductive accounts may be able to provide us with full-fledged elaborations of the neural underpinnings of those behaviors, but it isn't obvious that an answer couched in neurological terms is going to be always, and for every possible purpose, explanatorily satisfactory. We frequently demand explanations in folk psychological terms, regardless of whether we have reductive accounts of the terms being used. I don't think it is clear at all that every why-question we may raise in folk psychological terms is suitable to be satisfactorily answered in neurological terms. Thus, issues about irreducibility seem to be orthogonal to preoccupations about the theory's success.

---

(4) Sx *iff* P_1x or P_2x or … or P_nx

in which case the right side of the bi-conditional won't correspond to a natural-kind of neurophysiology. Ultimately, the reduced law that uses type-neurophysiologic predicates would look like

(5) P_1x or P_2x or … or $P_nx \rightarrow P'_1x$ or P'_2x or … or P'_nx

where Pi and P'i are nomologically related. The problem, however, is that if the identity relation in the bridge laws (like 4) isn't between natural-kinds, then they aren't laws. But if they aren't laws then (5) isn't a law either. And when no laws, no reduction. QED.

3.3 There May Not Be Beliefs After All

If you have been convinced by the considerations in the previous section, then you might think that the eliminative materialist does not have sound reasons for claiming that folk psychology is unsuccessful. In addition, if you consider the STA a valid argument, then you probably think that folk psychology is true. None of the above, however, gives you intentional realism yet. To that end, we still need one further argument, which may be called *the truth-to-existence-via-reference argument:*

(PP1) Folk psychology is true.
(PP2) The statements of folk psychology report propositional attitudes.
(PP3) Propositional attitudes are two-place relations between subjects and the referents of 'that'-clauses.
(PP4) All things considered, the best candidates we have for referents of 'that'-clauses are mental representations in the language of thought. Therefore,
(CC2) There are mental representations in the language of thought.

Again, each premise needs some clarification. (PP1) is the conclusion of *the success-to-truth argument* (i.e. (PP1) = (C1)). (PP2) is a traditional tenet that can be traced back at least to Russell's (1918) lectures on logical atomism. According to this claim, mental states are to be characterized as ascribing to a subject *S* an intentional verb *Vs* (such as 'believes', 'fears', 'hopes', etc.) and a certain proposition *p*. Propositional attitude reports, thus, conform to the following general form: '*S V*s that *p*', examples of which are "John hopes that it is raining", "Anne believes that having a small wedding is fine" and "Mario cree que el tiempo en Nueva York se siente distinto". Because propositional attitude reports conform to this general form, many believe that propositional attitudes are better understood as two-place relations between a subject and a proposition, which is the referent of the 'that'-clause. Indeed, it is customary to regiment propositional attitude statements in the following form:

[PA] $(\exists S)\ (\exists p)\ (R(S,p))$

where '*S*' refers to a subject, '*p*' refers to whatever the referent of the sentential complement clause may be (usually a proposition), and 'R' refers to the relevant intentional relation between them (Fodor 1978/1981; Schiffer 1992). Such is the rationale behind (PP3). In support of (PP3) Fodor gives three reasons[5] (Fodor 1978/1981: 178–179):

(a) "It is intuitively plausible. 'Believes' looks like a two-place relation, and it would be nice if our theory of belief permitted us to save appearances".[6]

[5] As mentioned, I'm confining my notion of intentional realism to Fodorian sentential realism. Because of that, the arguments in favor of (P3) and (P4) are his. Alternative accounts supporting (P3) and (P4) are not going tobe considered. It may be possible that my arguments apply to them as well, but they need not.

[6] Fodor uses "belief" as an illustration, but he's actually talking about all propositional attitudes. As such, his claims are to be read as extending to all propositional attitudes, not only to beliefs.

(b) "Existential Generalization applies to the syntactic objects of verbs of propositional attitudes; from 'John believes it's raining' we can infer 'John believes something' and 'there is something that John believes'."
(c) "The only known alternative to the view that verbs of propositional attitudes express relations is that they are (semantically) 'fused' with their objects, and that view would seem to be hopeless."

The force of all these reasons comes from linguistic and philosophical analysis of propositional attitude talk. The assumptions that support them will be discussed, when I present my arguments against (a), (b) and (c). Finally, (PP4) is basically an inference to the best explanation. The suggestion is that once you take into account all the data a theory of propositional attitudes is supposed to account for, the best candidate we end up with is a theory according to which "propositional attitudes are relations between organisms and formulae in an internal language; between organisms and internal sentences, as it were" (Fodor 1978/1981: 187). I think this inference to the best explanation can be blocked as well. Let us move on, then, to the challenges.

The first challenge goes against the claim, conveyed by (PP2) – and (a) – that mental states can (and need) be characterized as embedded within 'that'-clauses. It has been pointed out (e.g. Ben-Yami 1997) that some bona fide sentences reporting mental states cannot be rendered into the canonical form of propositional attitude reports ([PA] above). Consider the following sentences (examples 1 and 3, from Ben-Yami 1997: 85):

1. I want to sleep
2. Andrew knows how to multiply six digit numbers mentally
3. I trust Joan

A typical suggestion is to offer alternative paraphrases for these sentences, such as:

1*. I desire that I am asleep
2*. Andrew knows that to multiply six digit numbers mentally one needs to φ.
3*. I believe that Joan is trustworthy

But notice that these forced paraphrases introduce several problems. 1*, for instance, sounds odd. And this is not only a problem for English, as a quick look at the same proposition in French and Spanish, for instance, dissuades us from that option.[7] It may be argued that in order to get the correct, paraphrasing some extra linguistic maneuvering may be required, not at the surface level, but at the level of their deep structure (viz., 'that'-clause in 1 involves an implicit subject). Perhaps that could solve the problem for these cases, but if so one would like to know why we want to force our mental state reports to fit a certain kind of structure. I know of

[7] Contrast 1 with its Spanish translation "Quiero dormir" and its odd rendering into a canonical form: "Quiero que yo esté dormido". Ditto for French: "Je veux dormir" versus "Je veux que je sois endormi".

no argument to that effect (and neither does Ben-Yami 1997: 85). In the absence of such an argument it is hard not to conclude that the theory may be forcing the maneuver.

A related worry could be raised regarding 2*. I take it that all 2 tells us is that within Andrew's abilities we can count that of multiplying six digit numbers mentally. However, 2* seems to imply that if one were to ask Andrew how to multiply six digit numbers mentally he would be able to give us an answer in terms of φ. But 2* could be false while 2 be true. After all, Andrew may not know how it is that he manages to multiply six digit numbers in his mind. He knows that he can do it, but he may not know how or why he can do it.[8] And, finally, the same worry goes for 3*. All 3 tells us is that I trust Joan. It says nothing as to whether I believe that Joan is trustworthy. I could still stubbornly trust Joan despite the fact that I am seriously suspicious about her trustworthiness. Finally, I think that these considerations also speak against the first reason Fodor offers in support of (PP3). If not all mental states' attributions are suitable to be translated into statements of the canonical [PA] form, those that are can only constitute a subset of folk psychological statements. So it is not true that all folk psychological statements are better seen as two-place relations, as Fodor suggests.[9]

For the sake of the argument, however, let's assume that it is, in fact, intuitively plausible to render all our attribution of mental states in the canonical [PA] form. That is, suppose we accept that mental states can be paraphrased without semantic loss as expressing a two-place relation between subjects and the referent of 'that'-clauses – whether as propositions in abstracta or, as in the case of Fodor, presumably as neural concreta. Does that constitute enough reason to believe that the referents of 'that'-clauses are real? The answer is *no*. More assumptions need to get accepted for that conclusion to follow. Fodor gives us two reasons in support of (b): first, that 'that'-clauses behave referentially, and second, that existential generalization applies to 'that'-clauses. Now: why are these two reasons a good argument in support of there being referents of 'that'-clauses? It seems to me (and I'm not alone; see Balaguer 1998) that what underwrites this claim is basically Quine's criterion of ontological commitment plus an "intentional" reading of the Quine-Putnam indispensability thesis. Let me elaborate by comparing the case at hand with that of mathematics. Due to the influence of the Quine-Putnam indispensability thesis[10] in

[8] Notice that this is *not* a problem of expressibility. It isn't that Andrew does not know how to put into words what he does; it is rather that he may have no idea how he does it – he may not even know how to *begin* explaining what he does.

[9] A recent movement in epistemology, often called *intellectualism*, argues that know-how is a species of know-that (e.g., Stanley and Williamson 2001). If this was the case, then, it would follow that know-that statements should be translatable without semantic loss into know-how statements. Although arguing against intellectualism goes beyond the scope of the current essay, it may be worth pointing out that it remains a very controversial proposal, one that a growing number of philosophers reject (e.g., Noë 2005; see Fantl 2008, for a review).

[10] The claim, roughly, that if one's best scientific (physical) theory [after regimentation onto first-order logic] requires existential quantification over certain entities, then one is ontologically committed to such entities (Azzouni 1998: 1).

mathematics, theoretical irreducibility (and non-eliminability) is often assumed to carry with it ontological commitment. For it is frequently accepted that if *S* is irreducible to *R* ($=_{df}$ untranslatable to the other via bridge laws [see footnote 4]) and, when regimented, both *Sr* and *Rr* turn out to quantify over different variables,[11] then one is *eo ipso* committed to the existence of those entities (or kind of entities) picked up by the bound variables. In the case of mathematics such is the case with numbers (sets). I contend that for (b) to count as ontologically significant, the same should go for propositional attitudes (see also Balaguer 1998).

This argumentative line could be blocked with two moves. The first move is to show that 'that'-clauses do not behave referentially. The second move is to show that although existential generalization applies to 'that'-clauses, such a quantificational device can be read as being ontologically innocent, i.e. as conveying no ontological commitments by itself.

Let us begin with the first move. In general, objections against the non-referentiality of 'that'-clauses have been directed toward theories holding that the referents of 'that'-clauses are propositions. I believe that the force of at least two of these objections carry over to Fodor's analysis of propositional attitudes as being relational. The first of these objections in known as *the substitution failure*. Briefly stated the substitution failure objection says that if 'that'-clauses were really referential, and if their referents were really propositions, then they should share their denotations with linguistic constructions of the sort "the proposition that *p*" (Moltmann 2003: 82ff). However, this sort of substitution often fails. Consider the following substitution case:

4. John fears that Palin will be our next president.
5. John fears the proposition that Palin will be our next president.

Ex hipothesi, "that Palin will be our next president" and "the proposition that Palin will be our next president" share their reference: namely, the proposition that says that Palin will be our next president. But to be afraid of the eventual situation of Palin being the next president is different from being afraid of a proposition. It seems obvious that 4 and 5 differ in truth-value, so we should better conclude that 'that'-clauses do not refer to propositions (Hofweber 2006b). Now, does this concern carry over when we aren't talking about abstracta but concrete sentences in the language of thought? Consider:

6. John fears the mental sentence that Palin will be our next president.

Would 6 change the outcome of the substitution failure objection? I'm afraid not, at least insofar as the substitution failure objection counts as an argument *against* the relational analysis of propositional attitude reports. In order for (b) to count as a

[11] "Turn out" is short for: Take *Px* to be a formula with a free variable *x*, and take $\exists$ *(x)(Px)* to be directly deducible from S_r but not from R_r. Given Quine's criterion for ontological commitment, one is here committed to the existence of the referent of the variable in *Px* bound by the existential quantifier. Now: take $\exists$*(x)(Qx)* to be deducible from R_r but not from S_r. I take that if the criterion is correct, then it "turns out" that one is committed also to the existence of the referent of the variable in *Qx* bound by the quantifier (All under the assumption that one can have regimented versions of both *S* and *R*, my S_r and R_r Quine 1948).

linguistically valid reason in favor of 'that'-clauses being referential, Fodor needs that whatever goes for propositions goes too for mental formulae. And he cannot argue in favor of the latter as opposed to the former on the basis of some property that mental formulae but not propositions may possess. Remember that Fodor wants 'that'-clauses to be referential so he can claim, a priori, that there *must be* referents of 'that'-clauses. Using an alleged property about their nature to justify the argument in favor of their existence is circular.

The second objection I have in mind against 'that'-clauses being referential is originally due to Kripke (1979), although more recently has been developed by Bach (1997). The relational analysis of propositional attitudes finds support partly because it seems to reflect the apparent logical form of inferences like:

I1: A believes that *p*
B believes that *p*
→ There is something that both A and B believe.

However, when Kripke introduced his Paderewski-case puzzle he showed us that inferences of the form I1 aren't always valid. Suppose Carl meets Paderewski at a business meeting and as a result fixes the belief that Paderewski is a nice guy. Carl is pretty bad with faces, though. Later on he comes across Paderewski at a cocktail party where Paderewski strikes him as an annoying guy. As a result he forms the belief that Paderewski is not a nice guy. If the relational account of propositional attitude reports is correct, it seems as though Carl believes contradictory things. Specifically,

I2: Carl believes that Paderewski is a nice guy.
Carl disbelieves that Paderewski is a nice guy.
→ There is something that Carl both believes and disbelieves.

But Carl isn't being irrational; he's just ignorant about the fact that he's taking the name "Paderewski" to refer to two distinct individuals. Notice, however, that this fact is inessential to the problem. As Bach notes, when it comes to the relational analysis of propositional attitude reports, the believer need not have "any familiarity with the name in question or have any name at all for the object of belief" (Bach 1997: 224). Consequently, it seems that the two premises in I2 have Carl believing and disbelieving different things. If so, then I2 is not a valid inference. But given the fact that there aren't relevant formal differences between I1 and I2, we have no reason to believe that the linguistic appearances in I1 aren't misleading as well. To solve the puzzle Bach suggests that we reject an essential ingredient of the relational analysis of propositional attitude ascriptions: the assumption "that the 'that'-clause in a belief report specifies the thing that the believer must believe if the belief report is to be true" (Bach 1997: 221). In his account, 'that'-clauses *describe* their content instead (i.e., purport to state their content under a certain description, which may or may not be incomplete). Without this assumption, we have very little reason to take 'that'-clauses as referential.

Fodor can reject Bach's solution and stick to a relational analysis under the assumption that 'that'-clauses refer to mental sentences, which, unlike propositions, are neither ambiguous nor semantically incomplete. But this would be an unjustified move. Remember that (b) – and for that matter (PP3) – was supposed to convey pre-theoretical reasons in favor of 'that'-clauses being referential. Latching onto alleged properties of hypothesized mental sentences to save the linguistic phenomena whose clarity was supposed to motivate the relational analysis in the first place is question begging.[12]

Still, there is a further motivation to reject (b). Even if one accepts that 'that'-clauses are referential, the only reason Fodor seems to offer to jump from that linguistic fact to the conclusion that their referents exist is a commitment to an ontologically loaded reading of existential generalization. Since belief reports admit of existential generalization ranging over their 'that'-clauses (e.g., the example in I1), and since 'that'-clauses admit no reduction to another language whose ontological commitments we could be more comfortable with ("Behaviorists used to think such translations might be forthcoming, but they were wrong" [Fodor 1978], see also footnote 6), then we *should* go ahead, as Quine taught us, and accept the referents of 'that'-clauses as real (Quine 1948; see also Fodor 1987: 15).

Why would Fodor want us to do this? He cannot be suggesting this move on the basis of his acceptance of Quine's theory of reference; after all, Fodor is known for his rejection of Quine's holism tout court. A more plausible answer is that he is doing so on the basis of a weaker assumption: that the best – if not the only – way to understand existential generalization is by treating it as ranging over domain-independent entities. But this is a contentious claim. One can instead adopt what Hofweber calls "an internalist view" about quantification and deem existential generalization as a logical device to increase expressive power, and a logical tool that allows us to talk about infinitary disjunctions of single instances (Hofweber 2006a) – which is in this case, infinitary disjunctions of instances of attributions of mental states. If so, then, existential generalizations would be ontologically innocent.[13] The internalist view of existential generalization could turn out to be wrong, of course, but it is a good alternative. And without an argument against it – or without an argument in favor of a domain-independent reading of quantification – we would be better off remaining agnostic as to whether we should take existential generalizations as unquestioned carriers of the ontological burden of our regimented theories. As Jody Azzouni pointed out – in a rather different context – without an independent argument of that sort, it seems that the only reason we have to take the ordinary phrase "there is/are" to commit us to the existence of whatever it seems to commit us to, is simply "that the ordinary language 'there is' *already* carries ontological weight" (Azzouni 1998: 4). Does Fodor have an argument in favor of the reality of propositional attitudes independent of an ontologically loaded reading of existential quantification? He sure does – that's the bulk of the argument for (PP4).

Before we switch toward that discussion, however, let me say something very briefly about reason (c) for (PP3). In light of the previous considerations, it may be

[12] If we allow the resources of a theory to explain this phenomenon, a connectionist approach sensitive to graceful degradation and assignment by omission may turn out to do a better job than the language of thought when it comes to explaining why Carl forgot Paderewski's face to begin with.

[13] Free logic also allows to read existential quantifiers as ontologically innocent (Orenstein 1990).

clear that the force of (c) has now diminished. Fodor's original rejection of the "fusion" theory was supposed to mobilize the intuition that *unlike* that theory, a relational account of propositional attitudes faced no problems. But we have seen that relational accounts face severe objections too. Indeed, contemporary attempts to explain away precisely those objections seem to favor instead non-relational accounts of propositional attitude reports (see, e.g., Moltmann 2003, for a neo-Russellian account, as well as the appendix of that paper for other non-relational alternatives). Consequently, even if the fusion theory is false, we still need more reason to prefer a problematic relational account.

So what about (PP4)? Truth be told, Fodor can accept all the aforementioned objections and reject (PP3), and still argue in favor of his intentional realism on the grounds of (PP4) alone. He may say that, *all things considered*, intentional realism constitutes the best *empirical* theory we have to "vindicate" – his word – folk psychology. That is, he may well accept that we do not have either linguistic or a priori metaphysical reasons to accept the reality of sentence-like mental states, and still hold that such a hypothesis needs to be accepted on empirical grounds. At the end of the day, this has been his preferred strategy. Sheltered by the motto "the only game in town", the hypothesis of the language of thought has been advertised as the best theory we can muster to explain several psychological phenomena. Niceties aside, his argument boils down to an inference to the best explanation for some puzzling phenomena: concept acquisition, the compositional, systematic, and productive character of our thought, the projectability of mental terms in our psychological laws, and some (but not very many!) more. Copious pages have been written in an attempt to provide alternative accounts of these phenomena in terms that do not force us to accept a language of thought (see, for instance, Jackendoff 1992; Millikan 1984; Prinz 2002; Fodor 1990). I'm afraid I will not contribute to the discussion. Instead, I am going to try a different tack.

If Fodor's argument for the truth of intentional realism boils down to an inference to the best explanation, then it had better be the case that an inference to the best explanation constitutes a *good* reasoning pattern for realism about theoretical or unobservable entities. After all, folk psychology is just another theory – unrefined if you want, and operational over a slightly different domain than scientific psychology – but a theory none-the-less. Recall that folk psychology's mental terms are theoretical expressions whose alleged referents are unobservable inner episodes, i.e. mental states. Now, scientific realists usually take inferences to the best explanation as good argumentative patterns in favor of the truth of a certain theoretical hypothesis. In brief, the rationale behind the inference to the best explanation is that if a certain hypothesis *H* explains a certain phenomenon X better than any of its rival hypothesis, then *H*'s explanatory superiority should be taken as a mark of its truth – or, at least, as a mark of its approximate truth. From there, however, scientific realists often jump to the conclusion that the unobservable entities postulated by the theory must be real. Fodor, as we have seen, is no exception here. He takes the hypothesis of the language of thought to be the best hypothesis we have to account for the aforementioned psychological phenomena, and then goes on to claim that this is enough reason to believe that it is true that there are sentence-like representations in our brains.

Notwithstanding the widespread use of inferences to the best explanation by scientific realists, its validity as an argument to support the truth of a scientific hypothesis has been challenged on several grounds. Perhaps the most common attack comes from scientific anti-realism. To begin with, scientific anti-realists – like Bas van Fraassen (1980) and Nancy Cartwright (1983) – have argued that being a good hypothesis is never enough ground for believing that it is true. After all, the set of all rival hypotheses we can choose from may contain only false ones. Moreover, as van Fraassen remarked (1980: 21ff), when a scientist is in the business of accounting for some observational evidence, she does not really choose the best possible explanation *there is*, but rather the best explanation that is available to her. It would be a mistake to infer from that fact that such a hypothesis must be true, or closer to the truth than any other hypothesis she may or may not have access to.

Furthermore, van Fraassen also noted that most scientific realists take the thesis of scientific realism *itself* as an inference to the best explanation, insofar as it is the best hypothesis we can muster to explain the success of science (see Fine 1984). According to them, the success of a theory mustn't be cashed out in terms of sheer luck. Scientific realism is the best hypothesis we have to reject that preposterous conclusion. Now, the circularity of the maneuver isn't worrisome, yet it opens the door for a rival hypothesis to scientific realism, namely that "we are always willing to believe that the theory that best explains the evidence, is empirically adequate (that all the observable phenomena are as the theory says they are)" (van Fraassen 1980: 20). This anti-realist alternative to scientific realism, known as *constructive empiricism*, tells us that if a theory is successful, then it is empirically adequate, and that a theory is empirically adequate "exactly if what it says about the observable things and events in this world, is true – exactly if it 'saves the phenomena'" (van Fraassen 1980: 12).

My tactic to reject (PP4) should be obvious now; if Fodor's argument for intentional realism boils down to no more than an inference to the best explanation, and if inferences to the best explanation aren't conclusive reasons to believe in the reality of postulated entities, then (PP4) does not constitute a conclusive reason to infer the existence of mental formulae coded in our brains. With the previous arguments against (PP2) and (PP3), I tried to show that the jump from truth to existence *via* reference depended solely on the viability of inferences to the best explanations as valid arguments for the existence of unobservable entities. But as we just saw, inferences to the best explanation do not provide such conclusive grounds. Even if *all things considered* the language of thought turns out to be the best hypothesis we have to explain some behavioral (i.e. observational) phenomena, it is still unwarranted to infer that there *are* mental formulae in our brain. Again, I'm *not* saying that the hypothesis of the language of thought is false. All I'm saying is that the *truth-to-existence-via-reference argument* won't get us from the truth of our ascriptions of propositional attitudes to the reality of mental formulae in our brains. Which is why, I think, the best strategy for the metaphysically cautious philosopher of mind seeking to understand the place of propositional attitudes in our ontological repertoire is to approach the issue from an ontologically innocent anti-realist perspective (perhaps akin to constructive empiricism), and to proceed gradually, studying each propositional attitude ascription in its context of occurrence, the events – both behavioral and neural – with which they correlate, while taking as real only those parts of the explanations we have empirical evidence for.

3.4 Dennett's 'Prefutation' in C&C

To recap: In Sect. 3.1, I introduced the *success-to-truth argument* and suggested that both eliminative materialism and intentional realism spawned from different takes on it. In Sect. 3.2, I argued against Churchland's reasons to consider folk psychology unsuccessful. Finally, in Sect. 3.3, I presented some objections against the *truth-to-existence-via-reference argument* in order to prove it insufficient to support intentional realism. In the end I defended a metaphysically innocent approach toward propositional attitudes, very much in the spirit of van Fraassen's constructive empiricism, according to which our ontological commitments to the mental entities mentioned in our propositional attitude ascriptions should proceed in conformity with our empirical evidence in favor of their existence.

This is precisely Dennett's insight in *C&C*. He came to it from a different perspective, of course; he was arguing for the non-referentiality of mental terms and the plausibility of a fusion-view, according to which intentional statements should be taken as wholes when it comes to evaluating their truth values. However, his endorsement of the fusion-view was, at best, half-hearted. His real motivation, I believe, was to convince us that in order to advance the discussion about the reality of mental terms, we needed to temporarily withhold our grammatically driven metaphysical assumptions, at least until we reached a clearer understanding of the nature of the phenomenon whose reality is supposed to be at stake. His *tentative fusion* approach is, in this sense, methodological:

> We wish to proceed with no ontological presuppositions to the effect that mental entity terms either are or are not referential, and this can be accomplished by treating all sentences containing mental entity terms as tentatively fused, subject to further discoveries which will lead us to confirm the fusion or relax it. (*C&C*, 16)

Notice that the metaphysical innocence with which Dennett thinks intentional statements should be approached does not prevent him from regarding them as truth-evaluable:

> In most general terms our task is to provide a scientific explanation of the differences and similarities in what is the case in virtue of which different mental language sentences are true and false. Thus, for example, our task is not to identify Tom's thought of Spain with some physical state of his brain, but to pinpoint those conditions that can be relied upon to render the whole sentence 'Tom is thinking of Spain' true or false. This way of proceeding still characterizes the task of finding an explanation of the mind which is unified with, consistent with, indeed a part of science as a whole, but eschews—at least initially—the obligation to find among the things of science any referents for the terms in the mental vocabulary. (*C&C*, 18)

At this juncture, I think it is useful to see Dennett's view as a sort of re-interpretation of Sellars' myth of Jones in the spirit of van Fraassen's constructive empiricism.

Recall that, according to Sellars, back in the days of the mythical Jones, our Rylean ancestors were Positivists as well. They believed in a difference between observational and theoretical terms, according to which the former referred to observable entities and the latter to unobservable entities. But this dichotomy, as van Fraassen (1981) showed us, conflates two different distinctions: the distinction between observational and theoretical *terms*, on the one hand, and observable and unobservable *entities*, on the other. Whether or not an entity is observable has nothing to do with language: it has to do with observation. Accordingly, it is a mistake to think that because intentional terms got into our folk psychological language as theoretical terms, they must refer to entities that are unobservable, either in principle (e.g., states of the soul), or in practice (e.g., states of the brain).

Similarly, Dennett points out that the fact that our intentional terms appear to behave referentially does not necessarily mean that they must refer to some kind of unobservable entity, stuck in the middle of a causal chain of observable entities, and ontologically on par with them. Thus, he writes:

> So, one can only ascribe content to a neural event, state or structure when it is a link in a demonstrably appropriate chain between the afferent and the efferent. The content one ascribes to an event, state or structure is not, then, an extra feature that one discovers in it, a feature which, along with its other, extensionally characterized features, allows one to make predictions. Rather, the relation between Intentional descriptions of events, states or structures (as signals that carry certain messages or memory traces with certain contents) and extensional descriptions of them is one of further interpretation. […] The ideal picture, then, is of content being ascribed to structures, events and states in the brain on the basis of a determination of origins in stimulation and eventual appropriate behavioral effects, such ascriptions being essentially a heuristic overlay on the extensional theory rather than intervening variables of the theory. (*C&C*, 78–80)

Needless to say, the idea that we ascribe intentional states to others – as when we attribute propositional attitudes to them – as a heuristic to make sense of their behaviors (both afferent and efferent) became the pillar of what is oftentimes called the "instrumentalism" of the *intentional stance* (Dennett 1978, 1987). What I find surprising, having read *C&C* after studying much of what went on with the intentional stance in the 1980s and 1990s, is that critics typically accused Dennett of not respecting the ontological commitments that truth-bearing ascriptions of intentional statements, such as propositional attitudes, carry with them. To put it simply: critics thought that if he wanted propositional attitudes ascriptions to be truth-evaluable, then he had to take a stand regarding their reality. More precisely, critics thought that he either had to be committed to some sort of intentional realism if propositional attitude reports were to come out true, or some sort of eliminativism if they were to come out false. But Dennett didn't have to. He argued in *C&C* that whether a particular propositional attitude ascription comes out as true is independent of whether the intentional term embedded in it picks out something concrete in the brain (or in the soul). And this, I contend, amounts to a prefutation – i.e. a Dennettism meaning a refutation that is offered before an argument is raised (Dennett 1996) – of the claim that the truth (or falsehood) of our ascriptions of propositional attitudes carry ontological weight onto our theories about the nature of mental states – a widely shared but mistaken assumption in the realism/antirealism debate of the 1980s and 1990s about propositional attitudes.

I hope that this essay helps to place some arguments found in *C&C* within the context of the contemporary debate about truth ascription and ontology as it relates to intentional statements. No doubt there is much more that could be said about the relationship between truth ascriptions to intentional statements and the reality of propositional attitudes within Dennett's system. For instance, I think it might be worth exploring the extent to which the intentional stance can latch onto the theoretical resources offered by constructive empiricism when it comes to issues such as the reality of propositional attitudes. On the face of it, its seems like a relatively straightforward task. Traditionally, constructive empiricism and deflationism about truth have been lumped together. Given Dennett's Quinean inclinations it wouldn't be surprising if a constructive empiricist reading of his instrumentalism would end up supporting a deflationist view on the truth of propositional attitude ascriptions. However, recent developments suggest otherwise. As Jamin Asay (2009, 2012) has recently argued, constructive empiricism requires a more substantive theory of truth than deflationism. Would the same be the case for Dennett's view? In other words, does Dennett's instrumentalism require a more substantive view of truth than deflationism? If so, would that conflict with other Quinean aspects of his philosophy? And, what would be then the best truthmaking theory for Dennett's instrumentalism? These, I believe, are all questions worth asking, although their answers might have to wait for another day, and maybe for someone else.[14]

References

Asay, J. (2009). Constructive empiricism and deflationary truth. *Philosophy of Science, 76*(4), 423–443.

Asay, J. (2012). A truthmaking account of realism and anti-realism. *Pacific Philosophical Quarterly, 93*(3), 373–394.

Azzouni, J. (1998). On 'On what there is'. *Pacific Philosophical Quarterly, 79*, 1–18.

Bach, K. (1997). Do belief reports report beliefs? *Pacific Philosophical Quarterly, 78*, 215–241.

[14] A final, personal note: I read *C&C* for the first time in the summer of 2006. It was part of my background reading toward writing my MA thesis on the nature of propositional attitudes. I had read Dennett's work before, but never *C&C*. It also happened that, as soon as I finished part I of *C&C*, I went on to sail with Dennett and others on his boat *Xanthippe*, and at some point the subject of *C&C* emerged. 'Have you read it?' Dan asked. I told him that I had just finished the first part. 'And what did you think?' You see, at the time, I was not only working on my thesis; I was also working on my English, and my answer did not come across as intended. 'I was disappointed', I said, and laughter ensued. But what I meant to say is that I was disappointed to see that what I thought was an original idea in my MA thesis, turned out to have been there, masterfully articulated, in the first chapters of *C&C*! I did not abandon the project though, for notwithstanding the parallelisms between the claims in *C&C* and mine, I still thought it was worth showing how one could arrive at the same conclusion through a different path – this, I guess, is philosophy's way of reaching convergent evidence. Thus, the present essay draws heavily from my MA thesis, and I hope it helps to clarify my poor choice of words back when we were on *Xanthippe*! I also would like to thank the following people for their helpful comments on previous drafts: Jamin Asay, Jody Azzouni, Max Beninger, Alex DeForge, Dan Dennett, Anne Harris, Thomas Hofweber, Joshua Knobe, Gualtiero Piccinini, Jesse Prinz, and Kate Ritchie.

Balaguer, M. (1998). Attitudes without propositions. *Philosophy and Phenomenological Research, 58*(4), 805–826.

Ben-Yami, H. (1997). Against characterizing mental states as propositional attitudes. *The Philosophical Quarterly, 47*(186), 84–89.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Clarendon.

Churchland, P. M. (1979). *Scientific realism and the plasticity of mind*. Cambridge: Cambridge University Press.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, *78*(2). Reprinted in: Churchland, 1992, 1–22.

Churchland, P. M. (1985). Reduction, qualia, and direct introspection of brain states. *Journal of Philosophy*, *82*(1). Reprinted in: Churchland, 1992, 47–85.

Churchland, P. M. (1988). *Matter and consciousness*. Cambridge, MA: MIT Press.

Churchland, P. M. (1992). *A neurocomputational perspective*. Cambridge, MA: MIT Press.

Churchland, P. M., & Churchland, P. S. (1998). *On the contrary: Critical essays*. Cambridge: MIT Press.

Churchland, P. M., & Hooker, C. A. (1985). *Images of science: Essays on realism and empiricism*. Chicago: University of Chicago Press.

Clark, A. (1989). *Microcognition*. Cambridge, MA: MIT Press.

Cramer, J. G. (1988). An overview of the transactional interpretation. *International Journal of Theoretical Physics, 27*, 227.

Dennett, D. C. (1969). *Content and consciousness*. New York: Routledge.

Dennett, D. C. (1978). *Brainstorms*. Cambridge, MA: MIT Press.

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

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

Dennett, D. C. (1996). Did HAL commit murder? In D. G. Stork (Ed.), *Hal's legacy: 2001's computer as dream and reality*. Cambridge, MA: MIT Press.

Fantl, J. (2008). Knowing-how and knowing-that. *Philosophy Compass, 3*(3), 451–470.

Fine, A. (1984). The natural ontological attitude. In J. Leplin (Ed.), *Scientific realism*. Berkeley: University of California Press.

Fodor, J. (1974). Special sciences and the disunity of science as a working hypothesis. *Synthese, 28*, 97–115.

Fodor, J. (1978). Propositional attitudes. *The Monist, 64*(4), 501–524.

Fodor, J. (1981). *Representations*. Cambridge, MA: MIT Press.

Fodor, J. (1985). Fodor's guide to mental representation. *Mind*, Spring, 66–97.

Fodor, J. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

Fodor, J. (1990). *A theory of content and other essays*. Cambridge, MA: MIT Press.

Gallistel, C. R., & King, A. P. (2009). *Memory and the computational brain*. Chichester/Malden: Wiley.

Hempel, C. G. (1958). *The theoretician's dilemma: A study in the logic of theory construction* [Reprinted in: *Aspects of scientific explanation and other essays in the philosophy of science* (1965)]. New York: Free Press.

Hofweber, T. (2006a). Schiffer's new theory of propositions. *Philosophy and Phenomenological Research, 73*, 211–217.

Hofweber, T. (2006b). Inexpressible properties and propositions. In D. Zimmerman (Ed.), *Oxford studies in metaphysics* (Vol. 2). Oxford: Oxford University Press.

Horgan, T., & Woodward, J. (1985). Folk psychology is here to stay. *The Philosophical Review, 44*(2), 197–226.

Jackendoff, R. (1992). *Languages of the mind*. Cambridge, MA: MIT Press.

Kitcher, P. (2001). Real realism: The galilean strategy. *Philosophical Review, 110*(2), 151–197.

Kripke, S. (1979). A puzzle about belief. In A. Margalit (Ed.), *Meaning and use*. Dordrecht: D. Riedel.

Lange, M. (2002). Who's afraid of Ceteris-Paribus laws? Or: How I learned to stop worrying and love them. *Erkenntnis, 57*(3), 407–423.

Lycan, W. (2004). *Eliminativism*. (Unpublished) Available at: http://www.unc.edu/%7Eujanel/3255H5.htm

Matthews, R. (2010). *The measure of mind*. Oxford: Oxford University Press.

Millikan, R. G. (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.

Moltmann, F. (2003). Propositional attitudes without propositions. *Synthese, 135*(1), 77–118.

Noë, A. (2005). Against intellectualism. *Analysis, 65*(4), 278–290.

Orenstein, A. (1990). Is existence what existential quantification expresses? In R. B. Barrett & R. F. Gibson (Eds.), *Perspectives on quine* (pp. 245–270). Cambridge: Blackwell.

Piccinini, G. (2008). Computation without representation. *Philosophical Studies, 137*(2), 205–241.

Prinz, J. (2002). *Furnishing the Mind*. Cambridge, MA: MIT Press.

Quine, W. V. O. (1948). On what there is. *Review of Metaphysics, 2*, 21–38.

Russell, B. (1918). *The philosophy of logical atomism* [Reprinted in Pears, D. (1985)]. Chicago: Open Court.

Schiffer, S. (1992). Belief ascription. *The Journal of Philosophy, 89*, 499–521.

Schneider, S. (2011). *The language of thought: A new philosophical direction*. Cambridge, MA: MIT Press.

Schroeder, T. (2006). Propositional attitudes. *Philosophy Compass, 1*(1), 56–73.

Sellars, W. (1956). *Empiricism and the philosophy of mind* [Reprinted in: *Science, perception and reality* (1963)]. New York: Routledge & Kegan Paul Ltd.

Stanley, J., & Williamson, T. (2001). Knowing how. *Journal of Philosophy, 98*(8), 411–444.

Van Fraassen, B. (1980). *The scientific image*. Oxford: Oxford University Press.

van Fraassen, B. (1981). Critical study: Paul Churchland, scientific realism and the plasticity of mind. *Canadian Journal of Philosophy, 11*, 555–567.

Votsis, I. (2004). *The epistemological status of scientific theories: An investigation of the structural realist account*. Ph.D. dissertation, London School of Economics.