Attitude Ascription: Algorithmic or Artistic?

According to Frege, mankind has a common store of ways of thinking of objects and properties. Our thoughts are constituted by these common ways of thinking; we associate them with our words and sentences. We understand another's utterance when we know what way of thinking it expressed; to say what the other thinks, we find a sentence *S* which in our mouth expresses what she thinks and then say *the other thinks that S*. The early Russell had a similar view, minus the ways of thinking. According to Russell, there is a world of objects and properties about which we all talk and think; our thoughts are constituted by these objects and properties. To understand another is to know what proposition his utterance expressed; to say what the other thinks, we find a sentence S which in our mouth expresses what she said and use it to ascribe her attitude.

On these views, meaning is shared. And so the process of attitude ascription can be, as we might put it, algorithmic. Presented with an utterance, we decode it, finding the thought constituents and thought expressed. Then we find words that express the constituents and thus, when properly put together, express the attitude's object. And then we use those words to ascribe an attitude. An obvious worry, particularly about Frege's view, is that meaning *isn't* shared, at least not in a way that makes it reasonable to think that attitude ascription proceeds by content matching. As Frege himself observed, people think of objects in different ways. And as should be obvious once this is pointed out, by and large we don't have a clue, and don't care to have a clue, about *how* another thinks of the objects of which she thinks. When the other utters the sentence

Paracelsus was a chemist;

we do not pause to see if his store of Paracelsus information –or his conception of what makes one a chemist --resembles our own before saying

The other thinks that Paracelsus was a chemist.

Well and good, you say. This shows that Frege's picture of attitudes and their ascription was wrong. So let us dispense with sense and adopt Russell's view, on which what the other thinks is not made up of ways of thinking of Paracelsus and the set of chemists, but of Paracelsus himself and the property of being a chemist. Given that content is Russellian, there's every reason to think that it's shared, every reason to think that attitude ascription may proceed algorithmically.

Now, there are two reasons to worry about this response. The first is the well-known worry that our practice of attitude ascription is not well interpreted as one of ascribing Russellian content. We say, for example, that Mary may not know that Twain is Clemens, even though she knows that Twain is Twain. I want to concentrate for the moment on the converse worry, that (my use of) the ascription *the other thinks that S* may be true even though none of the other's thoughts has the Russellian content of (my use of) the sentence *S*.

Suppose I pretend to throw a ball and Diva the dog runs across the yard. Why did she do that? Because she thought that I threw a ball. The ascription is surely true. But the dog can't think what I say when I say 'I threw a ball' –that is, she can't think something with that Russellian content.

After all, the dog isn't disposed to group tennis balls, ping pong balls, footballs, and ball bearings together. But she would have to be able to do something like that to have the concept of a ball. And she would have to have the concept *ball*, were she able to think the Russellian content that I threw a ball. Else why think the dog thinks what I say with the sentence 'I threw a ball'?

The objection is *prima facie* incoherent. If it's *true* that the dog believes that S, how can it fail to be true that the dog believes what I say with the sentence *S*? If we have eliminated a Fregean account of the semantics of my sentence, then the semantic properties of the sentence I use are Russellian. Or they at least they are referential. When I use the sentence 'I threw a ball', the sentence has a part that refers to me, a part that refers to the property of throwing a ball. We may, of course, debate about what the property of throwing a ball is. It might be an abstract entity of some sort, or a function from worlds to sets of ball throwers, or just the set of this worldly throwers of balls. No matter: if the dog believes that I threw the ball, the dog has a belief that can be reported as having the content of my use of 'I threw the ball'. But that content is Russellian, or at least referential.

Note that to respond in this way is not to say that the dog has *our* concept *ball*, or that the dog has a representational ability that matches that of the English word 'ball'. Rather, it is to concede that the dog doesn't need anything much like our concept *ball* to think that balls are thus and so. Part of the point of the original objection is that the dog's believing that I threw a ball can't be a matter of its being in a state that *independently of and prior to my interpretation of it has the content of the sentence 'I threw a ball'*. This I concede.

Recall Putnam's discussion of the word 'water'. The way Putnam seeks to establish that 'meaning ain't in the head' involves three claims: the word 'water' 's reference in English in 2011 is not its reference in Twin English in 2011; the word 'water' 's reference in English in 2011 is its reference in English in the 18th century; the word 'water' 's reference in Twin English in 2011 is its reference in Twin English in the 18th century. The latter two claims tend to be treated as if they go without saying: after all, when an 18th century Englishman –Boswell or Dr. Johnson, say –uttered the sentence 'I want a glass of water', didn't he say that he wanted a glass of water?

Indeed he did. But to agree to this is not to say that the semantic properties of Boswell's utterance, independently of and prior to my interpretation of it, are identical or even all that close to those of the sentence I use to interpret Boswell's utterance. It is today determinate that 'water' is not true of more or less pure samples of D_2O, aka 'heavy water', which though visually indistinguishable from H_2O is lethal in large doses to fish and humans. It is hard to believe that this was *determinate* before the advent of modern chemistry.[1] The semantics of Boswell's words, independently of and prior to my interpreting them with my own words, are just not the same as the semantics that my words have, independently and prior to my using them to interpret Boswell. We sharpen the content of Boswell and Johnson's utterances when and by saying what they said in our idiom. In ascribing an attitude to them, we are not, or not merely, giving a report of a pre-existing identity of content. We are interpreting them.

The fact that Boswell's words do not, prior to our interpreting him, have the content we ascribe to him, does not mean that our ascription of that

[1] Here I echo thoughts of Joesph LaPorte. See the discussion in LaPorte 2004.

content is incorrect. Why shouldn't we say the same thing about the dog? Of course, this response raises a question: If the truth of my report *a believes that S* doesn't require a *match* of content between the sentence *S* as I use it and the content that some state of *a*'s has independently of my ascription, what *does* the truth of the ascription require?

Consider again the example of Boswell and the water. There is a certain semantic isomorphism between what is reported and what is used in the report. Boswell says 'I want a glass of water'; I report him with 'he said that he wanted a glass of water'. Ignoring some subtle issues connected with tense, the reporting sentence recapitulates the *term –verb –quantifier* structure of the vehicle of the attitude.[2] The report's accuracy turns on whether there is the "right sort" of relation between the parts of the attitude vehicle and the parts of the sentence in the report: each part of the sentence in the report has to be a "good interpretation", in a sense that needs specifying, of the corresponding part of the vehicle.

Considerations that have nothing to do with mismatches of referential content suggest that part of a correct account of attitude ascription will invoke the idea of the words in a report being a "good interpretation" of the vehicle of the attitude being reported. Consider a homely example.

> Eleanor Jane (=EJ) is Jane to her friends, Eleanor to others; her friends are well aware of this. You and I are her friends. I see that Bob and Ray, who know EJ but are not her friends, see her leave; I see that only Ray realized that it was EJ. (Of course, what Ray thinks

[2] Some think that 'wants' is actually taking a sentential complement in both vehicle and report. This doesn't affect the point.

> is 'there goes Eleanor.') I say to you "Bob and Ray saw Jane leave, but only Ray knew that it was Jane who left".

The way I speak in this example is quite natural. What information, exactly, would I normally convey by speaking in this way? How is it conveyed?

It is surely *not* conveyed that only Ray knew the Russellian content that Jane left. For Bob, knowing that *that* woman left, knew that too. Appeal to Fregean senses does not seem to be helpful –why think that the sense of my use of 'Jane' is a part of any thought known by Ray? Appeal to the idea that I am conversationally implying that Ray's belief is expressed by the sentence 'Jane just left' is no help, for it's part of the story that Bob and Ray don't know Jane is so-called.

It would be natural for me to speak as I do even if I didn't expect you to know whether Bob and Ray know EJ. So it would be natural for me to speak as I do even if I have no expectations about what you might assume about how they refer to EJ when they recognize her. What I convey in this example, roughly put, is that Bob and Ray saw Jane leave, but only Ray recognized her. That is: Ray had a bit of knowledge he might express with a sentence of the form *a just left*, where the sentence *That's a* would, for Ray, be good answer to the question, *Who's that?*; Bob had no such knowledge.

If that's what I convey, how exactly do I convey it? Here, the idea that attitude ascription requires that the words of the ascriber be a "good interpretation" of the representations of the ascribee can do some work.[3] Suppose that when we ascribe a belief saying *a believes that S*, we offer the sentence *S* as a **representation** or **translation** of what realizes one of a's

[3] What follows outlines the account I presented in Richard 1990.

beliefs. If so, attitude ascription presupposes something like a "translation manual", one keyed specifically to the individuals to whom attitudes are ascribed. In our story, I intend –and, if you understand me, you understand me to intend –that my use of 'Jane' in *Ray but not Bob knew that it was Jane who left* represents representations of Jane that identify her to the subject of the attitude. The semantic rule governing belief ascriptions is then something like this: an ascription *a believes that S* is true in a context just in case the sentence S, relative to context's translation rules, translates some belief-realizing state of a.

We may assume that translation requires preservation of referential content. But we may also assume that another's words or representations may, within limits, *acquire* content by being interpreted, as Boswell's word 'water' acquires the content of our word 'water' when we interpret his speech. Within a particular context, our interests can impose further, non-referential constraints on translation, typically keyed to particular individuals. When this happens, context contains rules of the form

> R1: When ascribing an attitude to person A, phrase *p* must represent aspects of A's states with property ϕ.

If, for example, we are concentrating on how the ancients might have expressed their astronomical beliefs, our context may presuppose the rule

> R2: When ascribing an attitude to the ancients, 'Hesperus' must represent representations that are associated with the ancient word which 'Hesperus' translates; ditto, for 'Phosphorus'.

Given the conventional account of what the ancients did and didn't know, this insures that while *Hesperus = Hesperus* translates an ancient belief, *Hesperus is Phosphorus* doesn't. So we speak truly when we say that the ancients believed that Hesperus was Hesperus, but didn't think Hesperus was Phosphorus.

In our story, Ray thinks to himself *That's Eleanor who just left*; Bob, on the other hand just thinks *that woman just left*. The operative translation manual is:

> R3: When talking about what Ray thinks, 'Jane' must represent Ray's customary representations of Jane.
> R4: When talking about what Bob thinks, 'Jane' must represents Bob's customary representations of Jane.
> R5: Otherwise, in representing their beliefs, we need only preserve referential values.

Given this, it's easy to see why it's true that Ray but not Bob realized that Jane left. Translating Ray's thought *Eleanor left* with *Jane left* doesn't violate the translation rules, as *Jane* is one of Ray's customary ways of thinking of EJ, and so the translation rules allow *Jane* to translate it. So it's true that Ray realizes that Jane left. But since Bob doesn't recognize Eleanor, there's no belief of his that can be translated with 'Jane left' without breaking a translation rule. So it's not true that Bob realizes that she left.

On this picture, attitude ascription is partially algorithmic –it involves something like a loose isomorphism of content. And it is partially artistic –it involves contextually variable decisions about how to render the simple bits and pieces of a thinker's thoughts. It will be said, however, that in the

general case, we can't hold on to even this much of the algorithmic picture of attitude ascription. Return to the example of Diva the dog. The semantic structure of the complement in the report *Diva the dog believes that I threw a ball* is *term-verb-quantifier*. Are we to suppose that the dog is in a state with an aspect or part that can be identified as a quantifier? Is the dog a candidate for a logic course?

I have mixed emotions about this challenge. On the one hand, I think we need to acknowledge that there is something out of whack about the epistemology suggested by the content matching picture of attitude ascription. Most anyone who watches me and Diva the dog knows that the dog thinks that I threw a ball. We know it without reflection and on the basis of no more evidence than our familiarity with the dog's fetching behavior and our current observation of the dog. But given the content matching account of attitude ascription, it can seem that in order for us to be justified in thinking that the dog thinks I threw a ball, we should be onto some fact that justifies our thinking that the dog has a mental structure that plays a role like that of the indefinite 'a ball'. But surely my observation of the dog justifies no such thing.

In the case of Bob and Ray, we have good reason to think that their beliefs involve representations of EJ and of the property of leaving the room –after all, we have excellent reason to think that those beliefs would be expressed by them with English sentences. In the case of Diva the dog, perhaps we don't have much reason to think that she is quantifying over balls, either those contextually available or otherwise. Still, we have reason to think that there is some state of the dog that is a belief and is a state that pictures the world accurately if and only if I threw a ball.

In fact, we have reason to think that the dog's state has *some* structure, even if it lacks the structure of the complement of

A. Diva the dog believes that I threw a ball.

As we usually think of that complement, it has a hierarchical structure, something along the lines of

A'. [[$_{DP}$ I] [$_{VP}$ threw [$_{DP}$ a ball]]].

Surely we have reason to think that the state that realizes Diva's thought has an element M that represents me. Don't we also have reason to think that there is some canine cognitive mechanism –TAB, call it --that is involved in the dog's belief and is reasonably interpreted as representing the property of throwing a ball? The dog, after all, is plausibly thought to recognize my action as being of a kind with your actual throwing of a ball, and as of a kind with young children's underhand ball tosses, etc., etc. She *does*, after all, reliably respond to all these in the same way. If the dog's ability to recognize this sort of action is implicated in her belief, then her belief involves something that is reasonably interpreted as a representation of the property of throwing a ball.

The state that realizes the dog's belief has *some* of the semantic structure of the sentence I use to report the belief. My sentence is a combination of the phrases 'I' and 'threw a ball', with the latter predicated of the former, so that the whole has a referential content that might be represented so

P. <Mark Richard, the property of having thrown a ball>.

The dog's belief state is composed of the representations M and TAB, with the latter predicated of the former. Thus, the dog's belief state has a referential content that is reasonably interpreted with my sentence.

When I think to myself 'she thinks I threw a ball', I do not have the sort of interests I have when I am focused on the ancients' beliefs about Venus or Bob and Ray's thoughts about EJ. So there are no special constraints, beyond the usual requirement of (loose) match of referential content, on how my words are to represent a belief state of Diva's. So in context the complement of (A) is indeed a 'translation' or representation of one of Diva's states of belief –'I' represents M and 'threw a ball' represents TAB.

I have been pointing out that understanding attitude ascription as involving a kind of translation doesn't require supposing that the belief states of others recapitulate all the semantic structure of the sentences with which we report their beliefs. It will be said that I haven't addressed the deepest problems with the idea that attitude ascription is a kind of translation. Consider an example of Dan Dennett's[4]. The way a computer plays chess may make it is correct to say that it thinks that it needs to get its queen out early. But there need be nothing in the machine that could plausibly be said to represent this belief *by* representing its parts. That the machine has this belief is not something that is true in virtue of its having and appropriately combining representations of itself, the temporal stages of a chess game, and a certain strategy that involves moving the queen up. After all. we would

[4] Dennett 1977.

say this sort of thing if the machine just has a pronounced tendency to move the queen out early, even when there's no immediate advantage in doing so.

Let us concede that the computer may have the belief though it has no state that realizes the belief by having proper parts that are adequately interpreted by proper parts of

B. One needs to get the queen out of early in a chess game.

Still, if the computer *has* the belief, it is in a state that realizes a belief and whose content is adequately rendered by B. If there are states that can have the content of B without recapitulating its structure, those states can be reported –and thus 'rendered' or 'translated' --using a sentence whose syntactic structure makes explicit the structure of the state's content.

It is no part of the ideas, that content is structured and attitude ascription is a kind of translation or representation, that belief states must be structured like sentences. A belief can be realized by a complex of dispositions –at least when those dispositions are embedded in the right sort of cognitive system –and dispositions lack sentential structure. Just as we may choose, in interpreting Dr. Johnson's utterance of 'water is tasty', to render it as involving the content of our term 'water', so we may choose, in interpreting the dispositions of the computer, to render them with our concepts of chess, the queen, and so forth. In the case of the computer, this is a good rendering because there is a rough parity between belief grounding dispositions of the computer and those that ground the belief of someone with an articulated belief about getting the queen out.

Some will say that beliefs realized by dispositions are properties of the whole organism. They will note that attitudes that are properties of the

whole organism need not be recorded by substates of the organism with semantic structure. And they will conclude that the contents of those attitudes are unstructured, and that ascription of those attitudes must be ascription of unstructured content. One argument to this conclusion goes as follows. It is determinate what the computer believes. But the determinacy in its belief does not extend below the level of truth conditions. After all, we can report the computer's belief with B or with

> C. It is generally a good thing for a chess player to move out her most powerful piece before the mid-board is crowded.

Each of these is acceptable because each captures what the world must be like to for the belief to be correct. There is, in the case of the computer, nothing more to capture about the content of its belief than its truth conditions; differences between B and C due to their semantic structures are irrelevant to saying what the computer believes. Thus, those differences must be irrelevant to capturing the content of its belief. Thus, sometimes an ascription of an attitude is an ascription of something whose content is unstructured; the content is simply the truth conditions of the attitude. But surely we are ascribing the same belief to the computer and to Sandrine when we say

> D. Both Sandrine and her opponent the computer think that a player should get its queen out of the back rank early in a chess game.

So even when there is a sentential record of a belief, that doesn't mean that the belief's content is structured. So the argument goes.[5]

It is true that we could use either C or D to ascribe a belief to the computer. It is true that the computer does not have an articulated representation of the semantic structure of these sentences. How does it follow –why is it even *plausible* --that differences between sentences that don't effect their possible worlds truth conditions are irrelevant to the content of the belief we report using them? It is not *just* the fact, that the computer's behavior is appropriate iff sentence B is true, that makes D true. It is also that the computer's dispositions are relevantly similar to the dispositions of people who have articulated beliefs realized by C. Like such people, the computer is disposed to manipulate certain objects (picked out in the sentences) in certain ways (also picked out in the sentences) at certain times. It is *this* difference between B and the complement of

> E. The computer believes that one needs to get one's queen out early in a chess game just in case the number of moves in the game will be equal to the product of some set of powers of primes

which explains why, thought B and E's complement are necessarily equivalent, B can be used to ascribe the computer a belief while E's complement cannot.

It is simply false that if a belief is realized by something that itself lacks semantic structure –a set of dispositions, say –then there is nothing more to the belief's content than what is given by the set of worlds in which it is true. The computer's belief about the queen is a belief *about* the queen

[5] Something like this argument can be found in the first chapter of Stalnaker 1984.

and the early parts of the game; it is not a belief about how these relate to arithmetical properties of the number of moves the game might contain. That this is so, of course, is manifest in what realizes the belief in the computer –the computer has dispositions to respond to and manipulate the queen in the early stages of the game; it presumably has none that relate number theoretic calculations of the likely length of the game to getting the queen out. There are more ways for the psychological facts about someone to make a structured content one that he believes then his simply tokening a sentence that expresses that content.

To have a belief is, *inter alia*, to represent a possible state of the world –to represent, that is, a particular distribution of objects, properties, and relations. The content of such a belief is a complex of objects, properties, and relations. To take the content of a state to be an unstructured set of worlds is in a sense to *deprive* it of content: Since contents are objects having properties and standing in relations, merely to ascribe truth conditions to something is not to ascribe a *determinate* content to it at all, but merely a range of possible contents.

So say I, but the fan of unstructured content is likely at this point to resurrect the above argument. The dog, it will be said, has a perfectly determinate belief. Since the dog's belief is determinate, it must have determinate content. But since the dog lacks a sophisticated representational system, it is not in a position to represent the structured contents associated with sentences we might use to ascribe it beliefs. But if the dog *determinately* believes a structured content, it must represent it –else why think that it believes *that* content, as opposed to one of the many structured contents that are equivalent to it? Why think that it is A that is determinately true, as opposed to

F. The dog believes that I propelled a ball through the air with a movement of my hand and arm.

As I see it, the question this argument raises is whether sentences like A or F can be determinately true when interpreted as ascribing a structured content to the dog. The answer is: Why not? Let us by all means agree that there is a good deal of slack between Diva's states, considered by themselves or in tandem with her everyday interactions with the world, and ascription of one or another concept of ball propulsion to her. Why should this lead us to say that A or F, if ascribing a structured content, is not determinately true? I would say that it is determinately true that Boswell and Johnson believed that water is wet. But I would not say that there is something about them and the relations they had to their environment and society in the 17th century that determines, prior to my interpretation, that the content of one of their beliefs is that *water* is wet. There is a certain amount of courtesy involved in the ascription of the belief that water is wet to Boswell and Johnson. That does not mean that the ascription is indeterminately true. Why is it that we cannot say the same thing about the dog?

Mark Richard
Philosophy, Harvard University
richard4@fas.harvard.edu

Dennett, Daniel. 1977. Critical Notice of Jerry Fodor, *The Language of Thought*. *Mind*.

LaPorte, Joseph. 2004. *Natural Kinds and Conceptual Change*. Cambridge University Press.

Richard, Mark. 1990. *Propositional Attitudes*. Cambridge University Press.

Stalnaker, Robert. 1984. *Inquiry*. MIT Press.