REPRESENTATION, TRUTH, REALISM

1. Many have noted the attractions of representationalist approaches in the philosophy of mind and in the philosophy of language – or should I say in the philosophy of mind and language. This essay is concerned with how some central issues in the debate over truth look when one sees them through representationalist eyes. The focus will be on telling the story as simply as possible without too much detailed argument but, of course, there will be some argument. Parts of the story will be familiar to some but I am sure that the story needs to be told.

I start by explaining the essentials of the representationalist framework. It leads naturally to an account of what it is for mental states which represent – beliefs, perceptions, and so on – to be true or correct or veridical, an account which counts as a kind of correspondence theory. At the same time, when extended from beliefs to sentences, the account explains why we say nothing new, in one good sense, when we append 'is true' to the sentence 'Snow is white' – the kind of point so often emphasised by opponents of correspondence theories of truth. The account also allows us to see why the case of propositions is very different: the reason for favouring a kind of correspondence theory for truth of beliefs and sentences does not carry over to truth for propositions. Finally, we note how the representationalist approach allows us to distinguish three kinds of realism.

2. Discussions of representation have a cast of usual illustrations. The

number of tree rings represents the age of a tree. The initial conditions of a deterministic system represent how it will be at any later point in time. Photos represent what they are photos of, maps what they are maps of, and diagrams what they are diagrams of. Sometimes it seems to be thought that what makes these cases of representation is their causal structure – in particular, their being one or another variety of causal co-variation. However, the relevant causal structures vary greatly and what unifies them is what the causal structures explain. What they explain is the systematic correspondences between the various ways some structure can be – the different number of rings, the different possible initial conditions, the different arrays of lines and shapes on paper – and the various ways something else can be[i] – the different ages the tree may have, the different states the deterministic system may be in at a given future time, the different ways what is photographed or mapped or diagrammed may be. It is the systematic correspondences that constitute the representation. Causation enters the picture because systematic correspondences of this kind would be sets of miracles in the absence of causal underpinnings.

What unifies our simple examples and makes them cases of representation is that each is a mapping from ways things might be to ways things might be, which we can symbolise as: $\{S_i\}\ M^j \text{—>} \{T_k\}$. All we require of the M^j is that it map from each S_i to one T_k. Thus, there will not be a single answer concerning what, say, S_{14} represents. What there will be is a single answer concerning what S_{14} represents relative to, say, M_3. There is no single answer as to what a barometer reading represents: air pressure, impending weather,

the internal state that is more immediately responsible for the reading, . . . ? There are many mappings from barometer readings to ways things might be, and correspondingly many right answers as to what the readings represent. What is true is that some are of much more interest than others, and all the ones of interest depend on causal connections in one way or another.

3. Some representations are correct, some incorrect. We are familiar with the map that represents the layout and number of stations in the London Underground and we trust that it does so correctly, but there is no *a priori* guarantee that it is correct. There might have been a glitch in the printing process, or perhaps a deliberate intervention by the bus company designed to discourage use of the Underground by tourists relying on the map. In either case, the map would misrepresent. The point above about relativity to mappings means that the difference between correct and incorrect representation is relative to a mapping. The map represents or misrepresents the layout and number of stations relative to such and such a mapping. When there is an especially salient mapping, as there is in the case of the London Underground – it is the one travellers on the Underground use to extract the information about the layout and number of stations by means of the map – we can talk without harm of representing correctly or incorrectly without further ado, but the relativity is there all the same. When a tourist says that they very much hope that the copy of the map they are using is correct, they mean correct relative to the mapping travellers use to extract information from the map.

We have, therefore, to define correctness relative to a mapping, not correctness *simpliciter*, and can do so thus:

> S_i represents correctly relative to M^j if and only if *a)* M^j maps S_i into T_k, and *b)* T_k obtains.

We can say this in terms of dividing possibilities into those consistent and those inconsistent with how things are being represented to be: to represent is to narrow the possibilities (down to those consistent with how things are being represented to be). There are many places the treasure might be and the role of a map is to cut them down to a manageable number. There are many times at which the next train might arrive and the job of a timetable to tell us which of the many is the actual one, the one which obtains. Of course, how things are being represented to be is typically indeterminate to one degree or another. The map does not say exactly where the treasure is and the timetable does not say exactly when the train will arrive. Also, the map and the timetable will be silent about a great deal: the temperature of the treasure, the weight of the train, the age of the universe are three obvious examples. We can capture indeterminacy and silence by spelling out the account of correctness in terms of sets of possible worlds, as follows:

> S_i represents correctly relative to M^j if and only if *a)* M^j maps S_i into a set of possible worlds, T_k, and *b)* T_k contains the actual world.

The variation among the members of T_k corresponds to the indeterminacy and

the silence. However, this strategy neglects an important property of a great many of the representations we rely on day to day. An important property of barometers is that they represent how things are with respect to themselves. The same goes for flags on ships, Morse code and for perception. A barometer represents how the air pressure is in its vicinity. A yellow flag on a ship represents that that very ship has cases of yellow fever. The SOS signal represents that there is an emergency where the signal comes from. (All relative to an appropriate, familiar mapping.) And perception represents how things are from the perceiver's point of view. For cases like these

> S_i represents correctly relative to M^j if and only if *a)* M^j maps S_i into a set of centred possible worlds, T_k, and *b)* T_k contains the actual centre and the actual world.

The actual centre for a given token flag state for the mapping function that captures the content of maritime flags is the ship flying the flag; *mutatis mutandis for other cases.* For our needs here, however, we can afford to neglect this important refinement and will frame our discussion in terms of divisions among possible worlds.[ii]

4. We can now say what truth is for belief. Belief is a representational state. To believe that snow is white is to be in a state that represents that snow is white. For many this is axiomatic. If belief isn't a representational state, what is? But here is a word or two of reinforcement. Diagrams and maps provide useful information about how things are; that's why we draw them and buy

them. The only credible explanation of how they provide useful information is that there are known mappings from the various ways they are to various ways the world might be. But that is to allow that they represent how things are: to stand in that kind of mapping relation is what it is to represent how things are. But the people who draw the diagrams and make the maps do so by using how they take things to be, their beliefs, together with their knowledge of the agreed mapping relations. It follows that belief represents – if it didn't, how could it serve to deliver what is needed to create the diagrams and maps – and that its representational content can be captured *inter alia* in those diagrams and maps. But, of course, it can also be captured in words: the informational function the map of the London Underground performs can be done with words, or with a mixture of words and the map as happens when tourists who need help reading the map ask at an information booth.

Truth for belief is a matter of representational correctness, as spelt out earlier, but we need to be careful how we fill in the detail. The basic picture is clear enough. Roughly, a belief represents that things are thus and so, and, in consequence, is correct or true if and only if things are as they are being represented to be. The need for care arises from the point that there are many representation relations, that representation is cheap. We said exactly this near the beginning when we said that representation is relative to a mapping M^i, which, of course, is why we had subsequently to define correctness relative to a mapping, not correctness *simpliciter*. But the example of belief allows us to reinforce this important point.

As good materialists, we know that token beliefs are neuroscientific states. Neuroscientific states casually co-vary with ambient temperature. Moreover, causal co-variance is agreed by all parties to be *a* representation relation. It follows that any belief token will represent its ambient temperature relative to one mapping that certainly counts as a mapping that makes for representation, but very few beliefs are about ambient temperature. In particular, a token belief that snow is white – my current belief that snow is white, as it might be – is not correct if and only if the ambient temperature is thus and so; it is correct if and only if snow is white. But, of course, the belief token does not represent the ambient temperature *qua* belief. There is a mapping that takes us from a belief token to how it, *qua* belief, represents things as being, to the set of worlds which are consistent with how the belief represents things as being. Call it the belief-mapping. The conditions under which the belief is true *qua* belief is that the belief is true if and only if the value the belief-mapping takes at the belief state in question is how things actually are, is the set of possible worlds that includes the actual world.

Similar formulations apply for other representational mental states like perception and desire. Perception represents that things are thus and so, and is veridical if and only if things are as the perception represents them as being. We may or may not accept that things are in fact thus and so. We may, in consequence, simultaneously be in a perceptual state that represents that the wall is blue while being in a belief state that represents that the wall is not blue (by virtue, say, of believing that the wall is white but is illuminated by blue light). It will not, though, be true in such a case that there is one mapping that

goes from the subject's states both to the wall's being white and to the wall's being not white. The mapping for belief will go to the wall's being white, whereas the different mapping for perception will go to the wall's being not white. A perception, then, is veridical if and only if the perception-mapping that goes from the perception to how it represents things as being takes an actual value at the perception, if it goes to a set of worlds that includes the actual world. Desire represents how one wants things to be, and is satisfied if and only if things are as the desire represents them to be; that is, if and only if the desire-mapping takes an actual value, a set of worlds that include the actual world, at the token desire state. And so on and so forth for hope, fear, regret and intentional states generally. For each, there is a distinctive mapping that goes from the state to its content, and the desire, hope and fear is realised if and only if the value taken at the state is how things actually are.

There are two general morals here. First (the one I have been labouring) is that we should not say that there is one representation relation and that the difference between belief, desire, perception, fear and so on is what stands in that relation – a belief, a desire, a perception, a fear or whatever. For each kind of representational mental state, there is a distinctive mapping that goes from the state to its content *qua* kind of mental state that it is, and the state is, respectively, correct, satisfied, veridical, realised or whatever if and only if the mapping in question goes from the state to how things in fact are. Indeed, for all we have said, a single token brain state might be both a belief and a desire, with a subject who is in that state counting as believing that *p* because the belief-mapping goes from the token state to that *p*, and counting as desiring

that *q* because the desire-mapping goes from the token state to that *q*. But having laboured the point, we can from now on afford to speak roughly and say that a belief is correct or true if and only if things are as it represents things as being. In using the term 'belief' for the state, we can think of ourselves in effect as selecting, from the many representation relations, the right one for belief. It is not unlike what happens when we declare that the morning star is a beautiful sight; what is meant is that it is a beautiful sight *in the morning*. This is why we may refrain from declaring that the evening star is a beautiful sight even when we know perfectly well that it *is* the morning star.

The second general moral is that truth for belief, veridicality for perception, satisfaction for desire, realisation for fear and so on are all of a piece; indeed, in English we can equally naturally use 'correct' for belief and perception alike, instead of 'truth' and 'veridical', respectively.

Finally, to avoid misunderstanding, I should note that treating belief as a representational state whose content is how it represents things to be is not to disagree with functionalist claims that content is a matter of functional role. Functionalism about belief should be thought of as an account of what we have called the belief-representation relation, of the mapping that delivers how a belief represents things to be. Here is one way the account might be spelt out (with details and qualifications suppressed in the interests of giving the core idea)

N belief-represents *S* if and only if i) *N* is a state designed to fit the

facts, ii) *N* moves its host in such a way that the host's desires tend to be satisfied if *S* obtains.

Functionalism is a theory of content in the sense of being a theory of the *determination* of content, not in any sense that competes with the representationalist picture of content as a set of possible worlds (or however we best represent the needed notion of how things are being represented to be).[iii]

5. The account given above of truth for belief is clearly a kind of correspondence theory, and the same goes *mutatis mutandis* for veridicality for perception, satisfaction for desire, and so on. A belief is one thing (a state of the brain according to we materialists); how it represents things to be is another. And the belief is true if and only if the second – something distinct from the belief itself, as we've just noted – is how things actually are. Of course, the metaphysics of what we are variously calling 'how things are being or might be represented to be', 'ways things might be', 'sets of possibilities', and 'sets of possible worlds' is very controversial. We discuss very briefly some views on the subject below. But, on no view on the market, is it the case that a belief (as opposed to its content) is the very same thing as how it represents things as being, and that is the key point which implies that we have here a kind of correspondence theory.

The next step in our story is to see how the observation that a kind of correspondence theory is correct for truth for belief leads to a kind of

correspondence theory being correct for truth for sentences, given an attractive position on the relation between belief and sentences.

The attractive position is that there is intimate connection between truth for belief and truth for sentences expressing belief. The conditions under which my belief that snow is white is true, and the conditions under which the sentence 'Snow is white' that I use to express this belief is true, are one and the same. Or, in the language of representation, how my belief that snow is white represents things as being, and how my sentence 'Snow is white' that I use to express that belief represent things as being is one and the same. Although there are many ways in which I can indicate how I take things to be – a shrug of the shoulders may signal that I doubt the veracity of what someone is saying; a pointing gesture may indicate where I believe the thimble is hidden – an especially good way for creatures who have mastered English (*mutatis mutandis* for other languages of comparable expressive power like French, Japanese *etc*.) is by means of a declarative sentence in English, and when that method is chosen the content of sentence and belief are one and the same.

How the obtaining of this intimate connection should be explained is a major topic in the philosophy of mind and language. Speaking for myself, I like the general approach to the question be found, for example, in John Locke, H. P. Grice, David Lewis and Jonathan Bennett.[iv] Mental content comes first and language is how we make public mental content through entering into arrangements to communicate how we take things to be to one another. On

this approach, the key clause (minus the many needed bells and whistles) that links truth for belief with truth for sentences might look roughly thus: declarative *S* is true in *L* if and only if i) the convention in *L* is to use *S* if and only if the user has such and such a belief, and ii) the belief in question is true. Inserting the earlier account of truth for belief gives: declarative *S* is true in *L* if and only if i) the convention in *L* is to use *S* if and only if the user has such and such a belief, ii) the belief in question represents *T*, and iii) *T* is the actual state of affairs.

But whether or not this account of why truth for belief and truth for sentences are intimately connected is along the right lines or not, provided there is the intimate connection, perhaps explained differently, a correspondence theory of truth for sentences will be correct. A (declarative) sentence will represent that things are thus and so in the way that the corresponding belief does, and accordingly counts as true if and only if things are as the sentence represents them to be.

A nice feature of this account is that it explains why the famous biconditional is *a priori,* consistently with a correspondence theory being correct for truth for sentences. Why is

'Snow is white' is true if and only if snow is white

a priori true? The explanation lies in the conjunction of two points: first, as representationalism tells us, the LHS is true if and only if the way things are is

as the sentence 'Snow is white' says they are, but, secondly, the sentence itself, the RHS, is a good way of saying what that way is. We can, that is, from the representationalist perspective see that the biconditional is true simply by knowing that the sentence named in the LHS is the sentence on the RHS, something that follows from our quotation mark conventions.

The role of these conventions is crucial. Consider English*, English with the one change that enclosing a sentence in inverted commas names the sentence enclosed with the first word replaced by the word 'water', so, for example, " 'Snow is white' " names 'Water is white'. In English*

'Snow is white' is true if and only if snow is white

is false, because it is false that water is white (although " 'Water is white' is true if and only if water is white" is *a priori* true, as English and English* agree about what gets named by enclosing 'Water is white' in inverted commas).

6. What about truth for propositions? For the representationalist this is an issue in metaphysics: the metaphysics of possibility, propositions and actuality. Consider the well-known views of Lewis on these three topics.[v] Modality is to be understood in terms of a realist metaphysics of possible worlds with world-bound individuals (each particular is in exactly one world) plus quantification. A proposition is a set of possible worlds. The actual world does not differ from any other world in having the property of being actual; the actual world for Jones is simply the world Jones is in; likewise for all of us; likewise for any

sentence token – the actual world for it is simply the world it is in.

On this metaphysics, a proposition is true if and only if the set of worlds which is that proposition contains the actual world, and whether or not this is the case is a question that can only be sensibly asked from the perspective of some individual or other, *I*, because being the actual world is being the actual world for *I*, the world where *I* is located. So if Jones asks what makes a true proposition true, the answer is that the proposition contains as a member the world she inhabits. Or take a metaphysics that agrees with Lewis's except for the extreme part of Lewis's realism about possible worlds – instead of an infinity of possible worlds each as concrete as the other, there is only one concrete possible world, ours, the actual one; the rest are abstract entities in some sense. On this metaphysics, the actual world is special in being the only non-abstract one, so if Jones asks the question what makes a proposition true, the answer is that the proposition contains as a member *the* non-abstract possible world.

A natural thought is that the most fundamental part of the debate over truth is over truth for propositions. As it is standardly put, a sentence is true if and only if the proposition it expresses is true, where this is understood as assigning conceptual priority to propositional truth. Or take our earlier rubric for sentential truth: a sentence is true if and only if things are as the sentence represents them to be, and the rough spelling out of this we gave above

> Declarative *S* is true in *L* if and only if i) the convention in *L* is to use *S*

if and only if the user has such and such a belief, ii) the belief in question represents *T*, and iii) *T* is the actual state of affairs.

A natural thought is that this passes the essential problem of truth across to clause iii), to the question of what it takes for a state of affairs to be the actual state of affairs. If we think of representation as a matter of making a division among possibilities (as we should, I suggested), then *T* will be a set of possible worlds in logical space – the set of worlds where the way things are in each is consistent with how the belief in question represents them to be, the set of worlds where the belief is true – and will be actual if and only if it contains the actual world. And what this in turn comes to will depend precisely on the metaphysical issues we reviewed in the immediately preceding paragraphs.

7. Finally we come to realism. For representationalists, there are three theses that can be classified as versions of realism. To explain what they are, I need to say a bit more about representation and divisions among possibilities. To represent is to divide the possibilities into those consistent with, and those inconsistent with, how things are being represented to be, as we said near the beginning in explaining how this handles indeterminacy and 'silence'. What then is the (representational) content of the belief, thought and saying that there are bears. It is the set of worlds where there are bears.[vi] But that set is not any old set: there is something that unites it, there is a commonality to that set, there is a pattern in those possible worlds. When we represent that things are thus and so – in thought or language – we are carving out similarity

regions in logical space. The diagram that represents the layout of the stations on the London underground picks out a certain configurational similarity concerning the layout of rail stations. The red colouring on maps that represent the onetime extent of the British Empire picks out a political similarity. The dot marked 'New York' on a map picks out the similarity of having something identical to New York (along with many other similarities like location, size, etc corresponding the essential richness of representation by maps) among the possible worlds where New York is as it is represented to be by that map.

Call these similarities patterns. Now we can state three realist theses as follows. The first realist thesis affirms that there are patterns to be found which obtain independently of us representers. The second realist thesis adds that we can we form conceptions of some of these patterns. Obviously some will be too complex, or too esoteric, for human minds to grasp, but the second realist thesis affirms that, all the same, we can grasp a great many of the patterns and think and say that *X* exemplifies the pattern. The third realist thesis affirms that as well as there being independent patterns which we can grasp, a good number of our opinions about where these patterns are to be found are justified and correct. Many beliefs and sayings to the effect that *X* is such and such are justified and true. The first thesis is a metaphysical one; the second is a conceptual one; and the third is an epistemological thesis. In terms of a simple example, the first thesis is that some things are square independently of our representing that they are; the second that we can represent that something is square in the sense that we know how we are

representing *X* to be when we think or say that it is square; and, the third, is that many of our opinions about whether or not something is square are justified and correct.

8. This is not the place to defend realism but let me conclude by noting the fallaciousness of an initially attractive argument for anti-realism about the metaphysics of patterns.

The argument runs as follows.

Premise. There are many commonalities in a sample and to say that one is *the* pattern exemplified is a mistake. (True)

Intermediate conclusion. It is a subjective matter which pattern is the one to cite in e.g. an IQ test. (True; pattern citing requires an implicit understanding of which of a number of candidates is the one to cite in a given context)

Overall conclusion. Patterns are a subjective matter.

The mistake in the argument is a confusion between the agreed subjectivity of which pattern it is right to select for some purpose, with the patterns that are candidates for selection being subjective.

Analogy. It is widely (though not universally) agreed that *the* cause of an event is a subjective matter in the sense that its being selected and named as

such is interest and context relative. There are many causes of a traffic accident – the relative locations and velocities of the cars involved; the many events that lead up to the cars being at those locations with those velocities, including the time the party started, decisions by traffic engineers, the power to weight ratios of the cars involved; and so on – and the rationale for citing one of the many events as *the* cause can only be an interest and context relative matter. But it does not follow from this that those many causes are a subjective interest and context relative matter. The subjectivity lies entirely in the assignment of the honorific 'the' to one or another of the many causes, not in being a cause *per se*.

Of course, emphasising the distinction between questions about patterns and questions about our selection of patterns for one or another purpose highlights the question that famously arises in discussion of the rule following problem, the question of how we succeed in picking out, for example, *plus* rather than *quus*.[vii] But that's another story.

Frank Jackson

The Australian National University

19 April 2005

---

[i] 'Something else' in the cases of interest. We can think of any physical item as representing how it itself is, and that, of course, will not be a matter of causal co-variation.

[ii] In my view, this important refinement undermines the moral often drawn from Twin Earth, see Frank Jackson, 'Narrow Content and

Representationalism – or Twin Earth Revisited', 2003 Patrick Romanell Lecture, *Proceedings American Philosophical Association*, 77, 2 (2003), 55-71.

[iii] For more on this theme see Robert C. Stalnaker, *Inquiry* (Cambridge, Mass., MIT Press, 1984).

[iv] See, e. g., H. P. Grice, "Meaning", *Philosophical Review*, 66 (1957), 377-88; David Lewis, *Convention* (Cambridge, Mass.: Harvard University Press, 1969); Jonathan Bennett, *Linguistic Behaviour* (Cambridge: Cambridge University Press, 1976); and John Locke, *An Essay Concerning Human Understanding*, Book III, Ch. II, § 2. The mention of convention in what follows is modelled on Lewis.

[v] David Lewis, *On the Plurality of Worlds* (Oxford: Basil Blackwell, 1986).

[vi] The switch from possibilities to possible worlds is in line with the previously announced policy of neglecting complications due to centering. It is arguable that to be a bear is in part to be a creature that stands in such and such a relation to *us*, in which case a full treatment would need to advert to divisions among centred worlds.

[vii] Saul Kripke, *Wittgenstein on Rules and Private Language* (Oxford: Basil Blackwell, 1982); the *plus* versus *quus* example makes its entrance on p. 8.