Mental Contents and Inferring

Devon Belcher

Mental Content and Inferring

One of the central tasks of philosophical psychology is to provide an account of the way agents revise and extend their store of beliefs. Rational agents extend their stock of beliefs not only through experience, but through logical inference as well. We reason from the beliefs we have to add new beliefs logically implied by them. Any theory that fails to accommodate at least the appearance of this ubiquitous phenomenon is seriously flawed. In this paper, I argue that such acts of inferring place important constraints on belief ascription and the theory of mental content.

In this paper, I focus on theories of content - contents of sentences, contents of beliefs, etc - which hold that such contents or propositions are identified with sets of possible worlds. This identification is itself an implementation of some broader view about minds and beliefs; such as the view that beliefs are partitions of logical space, or that beliefs are extensional dispositions to interact with the world, or whatever. Such a view of propositions is commonly referred to as coarse-grained because it individuates propositions so coarsely. For example, on such an account the proposition that Hesperus is Phosphorus cannot be individuated from the proposition that Water is H_2O. To deflate obvious objections, coarse-grained theorists often claim that what appears to be an ordinary belief about water is often a disguised belief about the word 'water' and what it refers to. We shall call this the metalinguistic rephrasal strategy.

My central argument is that the metalinguistic strategy, while it ameliorates some of the obvious problems faced by coarse-grained theorists, nonetheless fails to accommodate many mental phenomena. In particular, the metalinguistic rephrasal strategy, at least as commonly understood, cannot account for the truism that rational agents make valid inferences amongst

2

their beliefs (or that we can accurately report such inferences). In this chapter, I will introduce what I call the "argument from inferring", and apply it to the general metalinguistic rephrasal strategy. Although I will focus attention on Stalnaker's theory of belief attribution, the results of the argument are not limited to Stalnaker's theory alone. Generally, my arguments can be extended to apply to those theories that rephrase reports of necessarily false beliefs with reports of contingently false beliefs, whether the reconstrued beliefs are metalinguistic or otherwise[1].

A) Contents and Rephrasals

Our first question, then, is this: what is the content of a belief ? A family of theorists, including Carnap, Hintikka, Montague and (more recently) Robert Stalnaker, has long equated propositions with their truth conditions. These theories are often motivated by a position which holds that "belief and other cognitive states are [best explained] as capacities and dispositions to interact with the extra-linguistic world."[2] In cases where these extra-linguistic facts P and Q are necessarily equivalent (have the same truth conditions), a disposition to bring about P appears to be just a disposition to bring about Q. If this is so, then identifying the contents of beliefs with the truth conditions of those beliefs will ensure the identity of necessarily equivalent beliefs, as the dispositional account suggests. The formal device for representing this picture of belief is to treat the contents of the attitudes - propositions - as functions from possible worlds to truth values. Alternatively (and more or less equivalently), propositions are the sets of possible worlds which "make them true" . The proposition that the cat is on the mat, on this account, is the set of

---

[1]In the course of this paper, I hope to sketch some ways of extending my arguments to those theories which offer such treatments for the contents of assertions and beliefs as well as belief attributions. For simplicity's' sake, however, I here apply my arguments to those theories which rephrase the contents of belief attributions.

[2]Stalnaker, "Semantic for Belief" p 177, Phil. Topics 15 (1987)

possible worlds in which the cat is on the mat. Let's call this a coarse-grained account of propositions.

The immediate consequence of a coarse-grained account is that all necessarily equivalent propositions are identical. The proposition that the cat is on the mat is the same proposition as the proposition that the mat is under the cat. On the supposition that belief is a dyadic relation between an agent and a proposition[3], this theory leads to some very counterintuitive results. Suppose Sparky believes that recursiveness is recursiveness. This entails that Sparky believes that recursiveness is effective calculability. But surely one may truly report the former while truthfully denying the latter; after all, this identity was taken to be one of the great discoveries of the foundations of mathematics. Similarly, suppose Sparky believes that -(A v B). It seems that Sparky might fail to believe the logically equivalent proposition that (-A & -B); however, on the coarse-grained account, he is related by belief to one and the same proposition.

Moreover, the coarse-grained account forces us to attribute incoherent beliefs to otherwise rational cognitive agents. Since all necessarily false beliefs are identical, to believe one is to believe them all. Yet when we make an attribution such as

Sparky believes that Hesperus is Mars,

"we do not seem to be attributing to [Sparky] the necessarily equivalent belief that (for example) Hesperus is non-self-identical."[4] In order to avoid these results, many coarse-grained theories utilize a rephrasal strategy. In attributions of (apparently) necessarily false beliefs, the beliefs attributed are not really necessarily false propositions, but rather, related and contingently false propositions. These beliefs we attribute to an agent are actually beliefs about the relation of sentences to the propositions they (contingently) express. Such a rephrasal takes one of two forms. First, on the sentential metalinguistic rephrasal, attributions of the form:

3I restrict my arguments in this work to such a dyadic theory of the belief relation.
4Stalnaker, "Semantics for belief", p 179

2

A believes the proposition that Fa (where Fa is necessarily false)

are rephrased as attributions of the form:

A believes the proposition that the sentence 'Fa' expresses a truth[5].

On the other hand, on the sub-sentential metalinguistic rephrasal, they are rephrased as being of

the form:

A believes that the object referred to by 'a' is F.[6]

Let us call these the basic Metalinguistic Rephrasal Strategies[7]. Now the idea of rephrasing

problematic beliefs simply to avoid problems might appear *ad hoc* to some, especially if there

was no independent motivation for the rephrasal. But, as Stalnaker has pointed out, we do often

have independent motivation for rephrasal. For example, we can appeal to the principle of

charity, which tells us that we should always interpret a speaker on the assumption that he is

being rational. Clearly, (in most cases) to attribute a belief that Hesperus is not Hesperus to a

speaker is to treat that speaker as being irrational. The principle of charity suggests that we

rephrase the attribution in such a way as to attribute a coherent belief to the agent. Of course, this

new belief must be such that it is relevantly similar to the belief that the original attribution

---

5Alternatively, attributions of the form A believes that P are rephrased as being of the form
A believes that 'Fa' expresses the proposition that Fa.
6Again, an alternative is:
A believes that 'a' has the property referred to by the predicate 'F',
or even
A believes that the object referred to by 'a' has the property referred to by the predicate 'F',
depending on what one thinks the locus of A's confusion is.
7See, for example, Stalnaker 1987: "If I do not recognize some complicated truth-functional compound to be a tautology, and so doubt whether what it says is true, this can naturally be explained as doubt or error about what the sentence says." (emphasis mine)

seems to imply; and we should have a method for determining from a given utterance exactly what metalinguistic belief is *really* being expressed. The rephrasal strategies we are considering do just that.

Before we consider objections, there is an ontological point to make about each strategy. That is, each strategy discussed here is committed to the existence of heavily abstract objects, that is, propositions. These propositions contain or otherwise involve purely linguistic entities, that is, sentences. In this way, the strategies are unlike positivistic reductions such as (for example) Carnap's,[8] which reduces even propositions to purely physical objects, namely sentences (or at least sentence-types). On Carnap's account, belief is simply a relation (such as *accepting*) which an agent bears to a sentence. On the metalinguistic accounts we are currently considering, however, belief is irreducibly a relation to a proposition, and of course, in some cases (but not all, and in fact a distinct minority), those propositions involve sentences. Sometimes we believe that Hesperus is bright; sometimes we have beliefs about the ways that people represent their beliefs that hesperus is bright. The former belief is represented by a set of possible worlds in which Hesperus is bright; the latter belief is represented by a set of worlds in which people use a certain string of words (rather than another string of words) to represent the fact that Hesperus is bright.

Let us now consider the sub-sentential metalinguistic strategy. One can quickly see that changing the modal value of the proposition believed will create problems. Consider the attribution,

Kripke believes that water is H2O.

On the sub-sentential rephrasal strategy, this becomes something like:

8Carnap blah blah blah

Kripke believes that the substance referred to by the term 'water' is identical to the substance referred to by the term 'H2O'. (more colloquially, Kripke believes that the word 'water' refers to the same stuff as 'H2O'.)[9]

So far so good - on this account, Kripke believes a merely contingent fact about language (and if it is false, it is merely contingently false). Moreover, the set of worlds in which 'water' refers to the same stuff as 'H2O', is distinct from (for example) the set of worlds in which 'Hesperus' refers to the same thing as 'Phosphorus'. This gives the coarse-grained theory the intuitive result that one can believe that water is H20 without, say, also believing that Hesperus is Phosphorus.

Unfortunately, the sub-sentential strategy gives incorrect results when dealing with propositions involving embedded modal terms. Consider an example similar to the one just given. Kripke famously believed that all identities are necessary, and so it is also true that:

Kripke believes that necessarily water is H2O.

After all, this is a famous thesis in the philosophy of language, contrasted with earlier beliefs that water is only contingently H2O. But the sub-sentential rephrasal yields:

Kripke believes that necessarily, the substance referred to by the term 'water' is identical to the substance referred to by the term 'H2O'. (more colloquially, Kripke believes that necessarily, the word 'water' refers to the same stuff as 'H2O'.)[10]

This is surely a false attribution. As far as I am aware, very few people believe that words have

---

[9]Clearly, *the substance referred to by the term 'water'* must be read *de dicto* and not *de re*, or the content of belief is still a necessary truth, and rephrasal has gained us nothing.

[10]One might suppose that the correct rephrasal was really

Kripke believes that necessarily, the substance referred to by the term 'water' bears the relation expressed by the predicate 'is necessarily identical to' to the substance referred to by the term 'H2O'.

This added complexity seems unwarranted, however; for Kripke appears to understand perfectly the concept of necessary identity. The locus of his confusion is with water and H2O, not the relation they bear one another.

their referents necessarily; Kripke is not one of them. At any rate, one might surely have beliefs about the necessary identity of water and H2O without believing such radical claims about words having their referents necessarily. The sentential rephrasal, on the other hand, does not suffer this problem, for it yields simply:

Kripke believes that the sentence 'necessarily, water is H2O' expresses a truth.

Finally, the sub-sentential rephrasal strategy is unable to account for a large class of (seemingly) necessarily false beliefs. For consider an agent who understands quite well what the symbols '&', 'v', and so on, express (we may imagine that this agent has correctly answered questions about the associated truth tables, has given correct answers about the truth values of simple sentences involving these symbols, etc.). Such an agent might nonetheless erroneously claim that, for example, that the distribution principle:

A & (B v C) <-> (A& B) v (A & C)

is not a tautology. Intuitively, however, this cannot be explained as some error about what the individual words express, but rather about what the sentence as a whole expresses, in virtue of its structural / syntactic properties.

With these objections in mind, I hold that if the metalinguistic rephrasal strategy is to be adequate at all, we must use the sentential version. Now that we have outlined the basics of the coarse-grained theory of propositional content, and shown how it may defend itself against some obvious objections, I would like to take a brief detour and discuss some of the work that a theory of propositions is supposed to do. After we consider the role that propositions must play in any theory of our mental lives, we shall return to the coarse-grained theory of propositions and examine whether such a theory can be consistent with such an account of our mental lives. Not

surprisingly, the answer is 'no'- although the discussion of coarse grained propositions will grant us some general insights that we will apply over the course of this inquiry.

B) Propositions and inferring

One of the central tasks of philosophical psychology is to provide an account of the way agents revise and extend their stores of beliefs. Rational agents extend their stocks of beliefs not only through experience, but through logical inference as well. We reason from the beliefs we have to add new beliefs logically implied by them. Any theory that fails to accommodate at least the appearance of this ubiquitous phenomenon is seriously flawed.

To see what is we are interested in about inferring, consider Sparky, an ordinary person. Sparky is a rational agent and as such is constantly changing his mental states in light of new information. For our purposes, we may suppose that the relations between these changing states are causal relations (although these are likely to be no ordinary causal relations). Corresponding to many of these changing states are changing particular contents - or, as I will assume, propositions. And logical relations between contents are not plausibly causal relations. The proposition that the cat is on the mat does not cause the proposition (e.g.) that something is under the cat. Rather, these propositions are analytically connected. In certain distinctive cases, when there is an appropriate connection between these states, we say that Sparky has made an inference. All too often, Sparky (like the rest of us) does not change belief states in a way warranted by the logical or analytic connections amongst propositions. Sparky may change from a state of believing that the cat is on the mat to a state of believing that the cat is sleeping. In this case, Sparky has not inferred validly. That is, he has not made a logically valid inference. The inference is invalid because the contents of the belief states - the propositions Sparky believes

-fail to stand in the appropriate relation(s) to each other.

This suggests a necessary condition for validly inferring something from one's beliefs:

Agent A validly infers Q from his belief that P *only if*:
1) A believes P
2) A believes Q [11]
3) P entails Q (i.e.., P could not be true without Q being true).

Some comments are in order. First, note that this is not a sufficient condition. To be a sufficient condition, there would have to be an appropriate psychological relation between the states of believing P and Q - A's belief that Q would have to be brought about in some appropriate way by his belief that P. To hit Sparky on the head with a hammer on some occasion when he believes P, and thus bring about his belief that Q, would not be a case of Sparky's inferring Q from P.

Second, notice that the above condition holds only for inferring from one's beliefs. It is of course also possible to infer validly from information that is not actually believed. For example, one might *adopt as a premise* or *suppose* the proposition that P without actually believing that P, and subsequently infer from P the conclusion that Q. In so doing, one would come to have an associated propositional attitude towards the proposition that Q (e.g., accepting Q conditional on P). Evidently, a general necessary condition on valid acts of inferring which involve such attitudes would look much like the condition given. However, for simplicity's sake, I will restrict the arguments in this paper to cases of inferring from beliefs, and not acts of inference involving contents of other attitudes.

Third, notice that a similar constraint may hold for an attitude such as assertion. That is, it is plausible that agents sometimes (verbally) give arguments. This process may be thought of as

---

[11]Need (1) and (2) occur at the same time? I leave this open. Perhaps it is possible to infer P from Q in an instant; perhaps the process takes a short amount of time.

much like an act of inference. The agent asserts a premise, and concludes some proposition. If this is so, then a natural condition for giving a valid argument would be an adaptation of the above condition: for example, one would require that to give a valid argument one must assert the proposition which is the premise, assert the proposition which is the conclusion, and that the propositions asserted actually stand in the appropriate entailment relations.

Finally, notice that this condition implicitly treats belief as a dyadic relation. Nonetheless, there appears to be a analogous necessary condition for a valid act of inference even if belief were a triadic relation (such as a relation between an agent, a sentence, and the content of that sentence). If Sparky is related via sentence S to proposition P, and via sentence S' to proposition Q, then surely P must entail Q in order for any reasoning from P to Q (via S and S') to be considered valid.

Inferring and the Basic Metalinguistic Rephrasal Strategy

The central point of this paper is that the basic MLS cannot account for many ordinary acts of valid inferring. Suppose that one has correctly reported the following inferential act:

> Sparky validly infers that Phosphorus is dimmer than Hesperus from his belief that Hesperus is brighter than Phosphorus .

Such a true report is certainly possible, although the actual content of this report is subject to debate. However, we may use Stalnaker's own metalinguistic theory to find the content of the above report, and in conjunction with the necessary condition for valid inferring, demonstrate that this report, upon rephrasal, must result in an absurdity.

First, note that the report of the inference commits the reporter to holding that the necessary condition for such an inference has been met. That is, the reporter is committed to holding that

Sparky has certain beliefs; specifically, the beliefs that would be accurately reported by the attributions:

Sparky believes that Hesperus is brighter than Phosphorus (henceforth H>P).

and

Sparky believes that Phosphorus is dimmer than Hesperus (henceforth P<H).

Moreover, since this is a report of a valid inference, the reporter is committed to holding that the beliefs involved in the reported inference do, in fact, entail each other.

However, the beliefs ascribed to Sparky by these reports are necessarily false. Since Stalnaker holds that under ordinary circumstances (and this certainly appears to be such an ordinary circumstance) rational agents do not believe necessary falsehoods, these true reports must attribute some beliefs to Sparky other than the beliefs they appear to attribute {for if the reporter is attributing a valid inference to Sparky, then he must be committed to holding that it is the beliefs which Sparky does in fact have which stand in the appropriate entailment relations to each other}. The reports should be rephrased according to the MLS, using the very idiom of the original inference report. Thus, the rephrased beliefs actually involved in the report of inference are:

Sparky believes that the sentence 'H>P' expresses (in his native language) a true proposition[12].

and

Sparky believes that the sentence 'P<H' expresses (in his native language) a true

---

[12]Alternatively, the attribution might be that Sparky believes that the sentence 'H > P' expresses the proposition that H > Mars. The subsequent arguments go through mutatis mutandis in this case.

proposition.

Correspondingly, the original report must be rephrased to account for these beliefs involved in the reported inference. The rephrasal of the reported inference is:

> Sparky validly infers that the sentence 'P<H' (in his native language) expresses a truth, from his belief that the sentence 'H>P' (in his native language) expresses a truth.

But this is absurd. Our report of the act of valid inferring imputes an invalid inference to Sparky. Since, by hypothesis, we have correctly attributed inferences and beliefs to Sparky, it follows that Sparky has contingently false metalinguistic beliefs. But if this is so, Sparky will have made an invalid inference. We can see this by noticing that Sparky's act of inference has failed clause 3 of our necessary condition for acts of valid inference: that the proposition Sparky believes as his "premise" must entail the proposition Sparky concludes. This is not the case after rephrasal. Consider that in one world, '<' may not be the antonym of '>'. After all, it is a contingent fact that the symbol '>' expresses the brighter-than relation in the language Sparky happens to speak, and that '<' expresses the dimmer-than relation in this language. In this world, 'P > H' may express a truth yet 'P<H' express a falsity. In this case, it is not true that the concluded proposition is true in all the worlds where the premise proposition is true. Thus, Sparky's act of inference would be, *contra* our hypothesis, invalid.

D) Some possible responses

What responses are available? One might attempt to rephrase slightly differently so as to avoid this difficulty. For example, one might rephrase the original attribution of Sparky's belief in terms of language-relative sentences. For example, the attribution might be that Sparky believes that the sentence 'H>P'-in-English expresses the proposition that Hesperus is brighter

2

than Mars, where 'H>P'-in-English means something like 'the sentence 'H>P' in the language having the linguistic properties English actually has, where '>' expresses the brighter-than relation, and has the syntactical rules it actually has, etc.'. If the conclusion is similarly rephrased, then the premise initially believed would entail the premise concluded. This is so since, in all the worlds where 'H>P' expresses a truth, and '>' has the semantic properties it does (e.g., '>' expresses the contrary of '<'), 'P<H' will express a truth. But the content of Sparky's beliefs would then be necessarily false: there are no worlds (in which English words express the relations they do, etc.) where 'H>P'-in-English expresses a true proposition. Thus, the content of Sparky s metalinguistic belief - according to Stalnaker - requires rephrasal into some contingently false proposition, and the original problem arises again.

Another possible response is to treat the act of inferring as something like an enthymeme. On this response, an act of inferring like the above is an act with a suppressed premise. The suppressed premise would be that Sparky believes (perhaps tacitly) some semantic information required for the act of inference to be valid. In our above example, we might suppose that we have suppressed the premise:

> Sparky believes that the symbol '>' expresses the brighter-than relation; its contrary is expressed by the symbol '<';

With this premise taken into account, the act of inference is once again valid. In all conditions where 'H>P' expresses the proposition that H>P, and '>' expresses the contrary of '<' (and so forth), it will be the case that 'P<H' expresses the proposition that P<H.

However, this approach suffers from a problem similar to the first response. This problem is that in many cases, the added premise is incompatible with the initial rephrased premise. For example, suppose we report that

Sparky validly infers E(y)(x)(DimmerThan x,y) from his belief that E(x)(y)(BrighterThan x,y)

Again, this is an intuitively valid act of inference, although the propositions involved are, of course, necessarily false (since they entail that something is brighter / dimmer than itself).Thus, the attributions must be rephrased as:

Sparky believes that the sentence 'E(x)(y)(BrighterThan x,y)' expresses a truth

and:

Sparky believes that the sentence 'E(y)(x)(DimmerThan x,y)' expresses a truth.

As in the original example, the act of inference is no longer valid upon rephrasal. But according to the suggested response, this is simply because a vital premise has been suppressed. This premise might be (e.g.):

Sparky believes that the symbols 'E(x)', and 'A(x)' express the existential quantifier and universal quantifier, '<' and '>' express contraries, and those symbols have the syntactic properties they actually do.

In this case, Sparky's conclusion will be true in all the worlds where his premises are true. However, the premises are now incompatible. If Sparky believes that the quantifiers behave the way they do, then the conjunction of the two premises is necessarily false. There are no worlds in which the logical symbols express the relations and quantifiers they actually do, and the embedded sentence of the belief attributed to Sparky (as his premise) expresses a truth.

A final possibility is to apply the rephrasal strategy to the report of the inference itself, as well as to the beliefs involved in the inference. Such a rephrasal could take many forms. One possibility is to rephrase the original inference as:

Sparky validly infers* that Phosphorus is dimmer than Hesperus from his belief that Hesperus is brighter than Phosphorus .

2

-where inferring* is a relation with the following properties: for an agent to A validly infer* some belief Q from a belief P,

1) A believes P (where P is suitably rephrased)
2) A believes Q (where Q is suitably rephrased)
3) The proposition expressed by the embedded sentential complement of P entails the proposition expressed by the embedded sentential complement of Q.[13]

But this rephrasal appears completely unmotivated. Recall that the original rephrasal strategy was independently motivated by the principle of charity - that we should attempt to interpret our interlocutors in such a way as is consistent with their rationality. In this case, we have no such motivation. We have no reason to rephrase our interlocutors in such a way as all their inferences are valid. If a logic student, for example, inferred from A v B, that A, we would not be inclined to say he really inferred* A, where inferring* was valid when the proposition expressed by one embedded clause of a sentence (A, in the sentence 'A v B', in this case) entailed the other. Moreover, although we can make valid inferences* between necessarily false propositions on this account, the fact remains that we cannot make genuine valid inferences. We are presumably interested in the relations that hold between the contents of our beliefs when we speak of inferring, and not relations that hold between parts of our belief.

E) Conclusion

Doubtless, some of our logical errors are errors about what propositions are expressed by certain strings of symbols. But if all such logical errors are metalinguistic in *this* sense, then many of our intuitively valid acts of reasoning are not valid at all. I conclude that the basic

---

13Alternatively, the sentence formed by taking the embedded sentential clause of P, followed by the symbol '->' (or |-), followed by the embedded sentential complement of Q, expresses a truth.

MLRS is inadequate and cannot account for many of our (apparently) necessarily false beliefs or assertions. We can see from the examples in this chapter that any analysis of content which rephrases necessary truths as contingent ones will risk failure to preserve the relations amongst contents which allow for acts of valid inferring.

There is, however, one possible means of escaping this problem. If one supposes that words and sentences have their meanings necessarily, then the modal values of the rephrased necessarily true/ false beliefs are not shifted from necessary to contingent. And if the modal values remain the same, then phenomena that depend on modality - such as inferring - are not affected. Nonetheless, this approach does not appear available to Stalnaker for two reasons. First, suppose that words have their meanings essentially. Then rephrasing a belief that H>P to become a belief that

'H > P' expresses a truth

leaves the modal value of the sentence untouched; it is still necessarily false, and hence the same belief (on Stalnaker's account) as the belief that (e.g.)

Cats are really lobsters.

Second, suppose we can ignore the modal issues. Nonetheless, we have allowed certain entities into our ontology that are structured and abstract - namely, sentences. For I take it as intuitive that, no concrete entities could have their meanings essentially. So a sentence must be some sort of abstract entity; and such sentences appear to be at least partially individuated by their structure.[14] But now our ontology admits structured, abstract objects which can be processed by agents to yield the truth values of the propositions they express. Are these not more

---

[14]Stalnaker appears to endorse such a theory of sentences in blah blah, where he claims:

or less structured propositions, as the classical theorist has envisioned them? What remaining

reason is there to hold that propositions - qua objects of thought - have no structure?