Metaphor, Fiction and Thought

John Barnden [1]

Abstract. I will set out various un/underdeveloped opportunities for AI, philosophy and metaphor research to interact, with prospects for distinctly new lines of research and approaches to old problems. The opportunities I address in this paper are on the following topics: fiction-based accounts of metaphor, and a potentially resulting radical holism as regards the way metaphorical meaning arises from discourse; an anti-analogy-extension thesis, supporting unlimited non-parallelism between source and target in metaphor; the idea that thought can be metaphorical, and perhaps even more deeply than already mooted; deploying metaphor to solve a difficult problem in propositional attitude theory, which inludes the "meaning intention" problem as a special case; the "cognitive addition" of metaphor in language understanding, possibly leading to radical changes in how one thinks of the semantics even of non-metaphorical sentences.

1 INTRODUCTION

I will set out various un/underdeveloped opportunities for AI, philosophy and metaphor research to interact, with prospects for distinctly new lines of research and approaches to old problems. The opportunities I address in this paper are on the following topics, with the numbering corresponding to the sections of the paper.

2. Fiction-based accounts of metaphor, developed independently and under different names in various disciplines. One issue arising here is a possible radical holism as regards the way metaphorical meaning arises from discourse.
3. An anti-analogy-extension thesis, supporting unlimited *non*-parallelism between source and target in metaphor.
4. The idea that thought can be metaphorical, and perhaps even more deeply than already mooted.
5. Deploying metaphor to solve a difficult problem in propositional attitude theory (the problem being a generalization of the so-called "meaning intention" problem).
6. Something I call the cognitive addition of metaphor in language understanding, possibly leading to radical changes in how one thinks of the semantics even of non-metaphorical sentences.

There are threads strongly linking these topics. The dependencies will be summarized in the Conclusion section (section 7).

The paper draws heavily from already published papers and a journal paper under review (these will be cited below). In some places I incorporate partially-reworked extracts from those papers. However, the ideas have not all been drawn together before, or presented in a Computing and Philosophy venue, and some suggestions in sections 1 and 3 are new.

2 FICTION-BASED APPROACHES TO METAPHOR

I take a metaphorical expression such as "Ideas were whizzing around in his mind" to talk about a *target scenario* (here, a particular state of the mentioned person's mind and ideas) using the resources of a *source* subject matter (here physical objects and space).[2]

In various disciplines, researchers have suggested variants of an approach to metaphor that rests on what we can call *fictions*. Roughly and briefly, under such an approach the hearer of a metaphorical sentence uses the literal meaning of the sentence in context to (begin to) construct a fictional scenario expressed partly in source subject-matter terms. The fictional scenario is similar to a partial world as depicted by an ordinary fictional narrative such as a novel. The hearer may then elaborate (fill out) the fictional scenario by means of inference, using knowledge of the source subject matter. Metaphorical meaning arises when the hearer takes aspects of the fictional scenario and converts them into (alleged) aspects of the target scenario.

The fictional-scenario aspects that are so converted may either have been put there directly by the literal meaning of the metaphorical sentence, or may have arise through elaboration of the scenario. The created information about the target scenario forms part of the meaning of the sentence for the hearer. "Conversion" includes the case where an aspect is simply copied over to the target scenario without change, in the sense illustrated below.

This general characterization fits fiction-based approaches to metaphor in philosophy (see notably [42]), a recent enrichment of Relevance Theory accounts of metaphor developed in the field of linguistic pragmatics [18], and aspects of the "blending" or "conceptual integration" developed within cognitive science [21]. It is similar to the use of imaginary worlds for poetry understanding [31].

The characterization also fits the *ATT-Meta* approach to metaphor understanding that I have been developing and that is partially realized in a working computer program. I will describe this approach, as this will enable certain issues to arise in this section and other sections of this article.

The ATT-Meta approach makes an assumption that is contentious. Taking the above example of "Ideas were whizzing around in his mind," the approach does say that there can be a fiction in which an idea can do things like whizzing. Some may find this unintelligible. But perhaps this feeling can be allayed by the following. The approach in fact says that the stated whizzing implies that the ideas are indeed physical objects, in the fiction, as well as being ideas. In effect, the real-life fact that ideas are not physical objects is suppressed from becoming part of the fiction. (An important sector of the technicalities developed in the ATT-Meta computer program is for ensuring such suppression.) Another way of putting it is that it is presumably

---

[1] School of Computer Science, University of Birmingham, UK, email: j.a.barnden@cs.bham.ac.uk

[2] This statement is just a comment about metaphor, not a definition of it—and I am sceptical that it can be rigorously defined [7].

intelligible to state a counterfactual such as "If ideas were physical objects, then they could [do things like whizz around]." Fictions used in metaphor, at least according to the ATT-Meta approach, are akin to the bogus scenarios entertained in understanding counterfactuals. When compared to the real of ordinary fictions (novels, short stores, films, etc.) they are perhaps most akin to fanciful, fantasy tales.

However, a more moderate approach could have it that in the fiction there are physical objects that merely correspond to ideas outside the fictions rather than themselves also being ideas within the fiction. The rest of this paper would not be much disturbed by this alternative approach. (In fact, ATT-Meta probably needs to be extended to include the alternative anyway, in order to handle similes properly, such as in "The idea was like a balloon that was flying around the room." Currently, ATT-Meta would have to treat this in the same way as "The idea was a balloon that ...'.)

2.1   The ATT-Meta Approach

The ATT-Meta approach [2, 5, 6, 8, 12] is mainly geared towards cases involving familiar metaphorical views, such as the view of the mind as a physical region. However, the approach is not much concerned with conventional metaphorical phraseology based on such views, as in "The idea was at the back of his mind." Rather, it is geared towards open-ended forms of expression that transcend familiar metaphorical views. This is best brought out by examples such as the following:

1)   "The managers were getting cricks in their necks from talking up [to some people in power over them] and down [to the managers' subordinates]."[3]

It is common for abstract control relationships, especially in organizational settings, to be metaphorically viewed in terms of relative vertical position of the people concerned. However, someone having a crick in their neck is not a matter addressed by this view. Thus the sentence transcends the metaphorical view in question.

For purposes of (1), the fictional scenario is seeded with the premise that the managers literally got cricks in their necks from continually looking in two different physical directions, upwards and downwards to the mentioned sets of people.[4] This scenario gets elaborated, for instance by addition of propositions that the cricks cause the managers to have pain, emotional stress, difficulty in continuing such head-turning, and dislike of continuing it. These propositions follow just by ordinary commonsense knowledge about neck-cricks, etc. Some of these elaborated aspects of the fictional scenario get converted to become target-scenario propositions such as that (a) the managers experience annoyance and other emotional stress, and (b) it is difficult for the managers to continue the conversations.

Note especially that (1) does not just convey (a). The sentence is richer than if it had merely said that the managers were "getting annoyed" at their conversations. Annoyance does not in general imply difficulty of continuing, though it may imply reluctance to continue. However, in the fictional scenario, having a crick in their necks not only causes pain but also *makes it difficult for the managers to continue turning their heads,* and therefore difficult to continue the conversations. This difficulty is simply copied over to the target-scenario (by a mechanism to be mentioned below).

Example (1) and others have been analysed under the ATT-Meta approach (see for instance [3], [4], [6] and [9]). In example (1) the only "conversion" of fictional-scenario aspects into target-scenario ones were actually change-free transfers: difficulty of continuing the conversation in the fiction is converted to provide the same difficulty in the target scenario. But in general, genuine conversions are needed. This is illustrated by the following example:

2)   One part of Mary was insisting that Mick was adorable.

I take (2) to rest on two very general metaphorical views that are often used about the mind. First, there is the view of a person or a person's mind as having parts, where furthermore these parts are persons with their own mental states. I call these the "subpersons" of the person, and I call the view *Mind as Having Parts that are Persons.* (Note carefully that the parts are themselves a metaphorical fiction—the view *not* about objectively-existing parts of the person being metaphorically viewed as subpersons.) If a part (a subperson) of a person P believes (desires, intends, ...) X then, intuitively, the whole person P could be said to partly believe it. But what does it mean to partially believe something? The way I cast it is to say that the real person has a mere *tendency to* believe X. [5]

One main point of *Mind as Having Parts that are Persons* is that it allows different subpersons to have different beliefs or other types of mental state, and may even have beliefs that conflict with each other. This can rise explicitly in sentences that have a form such as "One part of P believes X, but another part believes Y" where X and Y conflict. In such a case the whole person P has tendencies to believe various conflicting things, without really *believing* any one of them. But I will also claim that the case of conflicting tendencies can arise implicitly, and in fact arises in (2).

The second metaphorical view comes into play when, as in in (2), the subpersons are portrayed as communicating in natural language. Since what is communicated is some idea that the whole person is entertaining, the additional metaphorical view here is that of *Ideas as Internal Utterances.* This is a very widely used metaphorical view that also often arises independently of *Mind as Having Parts that are Persons.* I will address the internal-utterances aspect of (2) shortly.

Now, there is a need to convert aspects of a fictional source scenario in which one or more "parts" of a person have particular mental states into aspects of the whole person's mental states in the target scenario. To handle fiction-to-target conversions, ATT-Meta borrows in part from conceptual metaphor theory (see [29], though more closely from [26]). A conceptual metaphor consists of a set of mappings—or as I will say, *correspondences*—between aspects of the source subject matter and aspects of the target subject-matter. These mappings constitute an analogy. The ATT-Meta approach broadly adopts this idea, though the correspondences are considerably different in form and function from those in conceptual metaphor theory and in analogy theory, as will be clarified below.

A metaphorical view in ATT-Meta involves a small number of very general, high-level, view-specific correspondences. In the case of *Mind as Having Parts that are Persons*, only two correspondences appear to be needed for a large array of examples. I just discuss one of them here. It can intuitively be expressed as follows.

(C)   A   person   *having   some   tendency*   to   believe/desire/intend/fear/like/... something corresponds metaphor-

---

[3] Cited in [25, p.162]. The example is from the *Daily Telegraph* newspaper.

[4] In discussing ATT-Meta previously I have usually used a weak notion of pretence rather than fiction, and have called the fictional scenario the pretence scenario. For present purposes the notion of a fiction is more revealing.

[5] Elsewhere I have cast this as the person having a "motive" to believe X, in a very general sense of a reason or some other factor. This is on the assumption that a tendency to believe something is underlain by a motive to believe it. Here I revert to an earlier, more theoretically neutral formulation in terms of tendencies.

ically to *at least one subperson* of that person having a tendency to (respectively) believe/desire/intend/fear/like/... it.

C can be deployed by the hearer of (2) as follows. Taking sentence (2) literally, the hearer puts the premise that (literally) the mentioned part of Mary insists that Mick is adorable into the fictional scenario. This fictional claim is used to infer that (by default) the part is a subperson inside Mary. Given the general default that when people claim things they believe them, the hearer can then infer that, still in the fictional scenario, *that subperson believes that Mick is adorable.* It follows *a fortiori* that that subperson has a *tendency* to believe that Mick is adorable. Then hearer converts that fictional-scenario claim using (C), to become the target-scenario claim that *Mary* has some tendency to believe that Mick is adorable.

But also the insistence in (2) can be used to infer within the fiction that actually there is a subperson of Mary that believes that Mick is *not* adorable. This is because of the real-world nature of insistence. Typically, someone insists something when there is a conversation with a person who denies it. Thus, the presence of a subperson who claims that Mick is not adorable can be inferred by default. This new subperson presumably *believes* that Mick is not adorable. Hence, again using (C), we conclude that Mary has a tendency to believe that Mick is *not* adorable, as well a tendency to believe that he is.[6]

A final comment on (2) is that it crucially involves the notion of insistence by fictional subpersons, but this notion does not need to have its own correspondence to any non-metaphorical notion about the person's (Mary's) mental states. In short, insistence as such does not need to be handled by any correspondence associated with the two metaphorical views mentioned above. The insistence was used merely to generate, within the fictional scenario, certain conclusions that could be mapped by (C). If insistence does not have its own tailor-made correspondence associated with any metaphorical view the hearer knows, it is a view-transcending aspect of (2).

However, assuming that an utterance by a subperson is (metaphorically speaking) an utterance inside Mary, and assuming that *Ideas as Internal Utterances* involves a mapping of such utterances to thoughts of Mary's, then there is an additional line of processing leading to conclusions that Mary is entertaining certain thoughts.

One difference between ATT-Meta's approach and (other forms of) conceptual metaphor theory is that in ATT-Meta there are two broad sorts of correspondence: (i) *view-specific* correspondences such as (C), associated with particular metaphorical views, and (ii) *view-neutral mapping adjuncts* that apply by default in any case of metaphorical understanding, irrespective of what metaphorical views are in play, and that build upon the effects of, and indefinitely extend the reach of, the view-specific correspondences. Returning to the neck-crick example, (1), how can the hearer create target-scenario conclusions such as that the managers, in the target scenario, experience negative emotions, caused by the conversations, and find it difficult to continue their conversations? Such conclusions arise within the fiction, but they need to be transferred to the target scenario. The crucial observation here is that there are general qualities about metaphors' fictional scenarios that are very often copied in metaphor to the target scenarios no matter what the specific metaphorical view is. Amongst such qualities are the following:

- Emotional/attitudinal states, value-judgments, etc. (of typical ob-

servers such as the hearer to the target scenario, or of agents within the scenario itself).
- Mental states, such as believing, intending, wanting.
- Time-Course, incl. starting, continuing, ending, immediacy, smoothness/intermittency, rates at which episodes occur, temporal relationships between episodes, etc.
- Causation, prevention, enablement, ability, attempting and tendency relationships, and related qualities such as effectiveness.
- Ease/difficulty properties.

For each of these qualities there is a *View-Neutral Mapping Adjunct* (VNMA) that allows transference of aspects of a suitable fictional scenario to the target scenario. In our neck-crick example, one VNMA delivers a correspondence between emotional distress of the managers about the conversations, in the fiction, and emotional distress of the managers about the conversations, in the target scenario. The VNMA concerned with causation allows the inference that the fact that the conversations *cause* the emotional distress in the fiction is inferred to correspond to their also doing so in the target scenario. Equally, the within-fiction difficulty for the managers of continuing with the conversations transfers to the target scenario, because of VNMAs handling time-course (a case of which is the continuation of a situation) and difficulty. The continuation of a situation is one case of a qualitative temporal attribute.

While (1) only involves the use of VNMAs and (2) uses only view-specific correspondences, both types of conversion mechanism are needed in general. Both types are defeasible, so their results can be defeated in specific circumstances by other evidence.

One important facility currently missing from ATT-Meta is an ability to discover novel analogy between two scenarios. In a minority of cases of metaphor, and quite often with cases of so-called image metaphor (resting largely on physical appearance), there are no existing correspondences that will deliver useful results. However, a novel-facility could readily be added without disturbing the existing nature of the approach.

2.2 Issues for Fiction-Based Approaches

By virtue partly of having been realized in a working computer program, it is fair to say the inference and conversion mechanisms in ATT-Meta have been worked out much more specifically and completely than in fiction-based approaches developed in non-computational research endeavours, even though much more work needs to be done on ATT-Meta itself (both theory and program). The work of computationally operationalizing fiction-based theory has thrown some general issues into relief, all of which I believe need further research and, more particularly, could benefit from collaborative research between philosophy, metaphor theory and AI.

First, it is not rare for ordinary fictional narratives to meld several entities, such as people or places, in the real world into a composite entity in the fictional world. Ordinary fictional narrative can also do the reverse, i.e. have several different entities in the fictional world correspond to one entity in the real world. Such violations of one-to-one mapping between fiction and what lies outside the fiction raise philosophical issues—e.g., about the nature of fictional entities and about cross-world correspondences more generally—and detailed computational issues as regards representation and inference, while also possibly being important in metaphor. However, they have been little studied in the metaphor area. This may be partly because they are rare in metaphor—but the matter has not seen much explicit exploration. That it may not be rare is suggested by the *Mind as*

---

[6] As pointed out by a reviewer, (2) suggests that Mary is actually having conscious, occurrent thoughts about Mike. This addition to the interpretation of (2) can be handled by assuming that (C) covers such thoughts, and recognizing that when someone claims something X, insistently or otherwise, they have a conscious, occurrent thought that X.

Having Parts that are Persons view. Although ATT-Meta does not currently in fact postulate a mapping between the actual person and the fictional subpersons (as opposed to the above partial correspondence (C) between the mental states of the actual person and those of the subpersons), this might be a valid basis for analysis. Conversely, utterances such as "The country wants to abolish slavery," when analysed as metaphorical, could perhaps be cast as metaphor that puts one thinking agent in the fiction (that agent being the country) in correspondence with a large number of thinking agents in the country.[7]

Notice here in passing that, again, an element of the target scenario can also appear in the fictional source scenario, either with merely its properties from the target scenario or with a partially different set of properties. The country in the slavery example just mentioned is in both the target scenario and the source scenario, but in the latter it is a thinking agent as well as a country. We saw an analogous phenomenon when discussing ideas whizzing around in someone's mind: the ideas were in the source scenario as well as the target scenario, but in the source scenario they were physical objects as well as ideas. This use by a fiction of elements from outside it, with possibly a warping of the nature of those elements, is familiar from ordinary fictional stories.

Secondly, I have argued elsewhere [11] that metaphor understanding can be facilitated by "reverse" conversion steps, i.e. ones in the target-to-fictional-scenario direction, as well as ones in the normal, forwards direction. Such reverse conversion is in fact implemented as standard in the ATT-Meta system. The most interesting basis for wanting reverse conversion is a claim that it is sometimes easier to find coherence between related metaphorical utterances in a discourse and surrounding or interspersed utterances by looking to the fictional scenario rather than to what the fictional scenario says about the target scenario. Reverse conversion brings fiction-based theory of metaphor closer to the theory of fiction in general, given that it is standard for ordinary stories to bring in information about the real world. For instance, if we know that a certain fictional character is intended to correspond to a real person, we would tend to import our knowledge of that person into the fiction (if not contradicted there) suitably amending it to fit the circumstances of the fiction. Yet reverse conversion is not extensively considered in metaphor research. (It has been mooted without extensive detail in the context of Interaction theories of metaphor [41], and has been discussed in some applications of the blending approach)

Thirdly, I have also argued elsewhere (e.g., in [13]) that a metaphorical sentence sometimes cannot readily be given its own meaning in terms of the target scenario. Rather, it may conspire with surrounding literal or metaphorical sentences to convey something about the target. This is a form of holism about discourse meaning. The general point is that several sentences in a discourse might need to contribute to building up a fictional scenario (perhaps with the help of reverse conversion, if literal sentences are involved) and to allow appropriate elaborations that lead to fruitful opportunities for fiction-to-target conversion. However, following traditional assumptions about literal sentences , language researchers in many disciplines appear to assume virtually without argument that every sentence, including metaphorical ones, must be assigned its own meaning in terms of the situation actually being talked about. However, I conjecture that it is merely a *typical* case that a sentence taken alone

can be assigned such a meaning. Rather, meaning can act much more holistically across sentence (or clause) boundaries, and there is no hard syntactic limit as to what sort of segment of discourse might in a particular case be treated most naturally as a unit bearing specific meaning.

An example I use in [13] is

3) "Everyone is a moon, and has a dark side which he never shows to anybody." [attributed to Mark Twain by [17, p.74]]

Note that the example could just as well have been in the following multi-sentence form, which is just as comprehensible:

3a) "Everyone is a moon. Everyone has a dark side which he never shows to anybody."

I suggest that it is misguided to suppose we must first derive a metaphorical meaning for the clause/sentence "Everyone is a moon" and a metaphorical meaning for the clause/sentence "[Everyone] has a dark side which he never shows to anybody" and then combine these meanings. Rather, the second clause indicates what it is about being a "moon" that we should attend to (this isn't provided by the first clause), while it is the first clause that brings moons into the picture (the second clause doesn't do this). I claim the best approach is to form a fictional scenario on the basis of both clauses, and only then extract implications for the target scenario. In the fiction, the moon aspect reinforces the never-showing aspect of the second clause.[8]

Now, the second clause in (3) or second sentence in (3a) could plausibly have been given a metaphorical meaning even if the first clause/sentence hadn't been uttered. The fiction would have just cast the person as *some* physical object that has a dark side not shown to anyone else. So, for (3/3a) itself, one can imagine a process whereby the hearer works out that metaphorical meaning for the second clause/sentence and only later refines or strengthens it in some way by means of the first clause/sentence.

But the main point I wish to make is that it would be quite hard to give the first clause its own metaphorical meaning, and therefore quite hard to form an integrated understanding by taking a metaphorical meaning for the sentence and a metaphorical meaning for the second and combining them. . Either it would involve using the second clause for guidance as to what the first one means, in which case there hardly seems any point considering the first clause at all by itself, or the operation would involve taking the clause in isolation of the second, in which case (unless surrounding discourse context could help) we have the usual problem of the indeterminacy of metaphor (see, e.g.,[39]). Without the second clause it is wide open what the first clause is getting at. For example, it could be construed as saying that everyone is somehow subservient to something that is being metaphorically portrayed as the Earth, or as saying that everyone serves as a source of illumination for the world in times of darkness, or ...

Actually, the first clause has a deeper effect than just reinforcing the never-showing in the second clause. The moon also has a bright side, at least some of which we can normally see, and which is extremely salient in a clear night sky. Thus, a more elaborated interpretation of (3) or (3a) could include the notion that everyone also has a side that is (in part) usually very much apparent. This new message cannot come from just the second clause, because although the mention of a dark side weakly suggests a non-dark side, there

---

[7] Sentences such as "The country wants to abolish slavery" would typically be analysed as involving a *metonymic* step from country to (some/many) people in the country. But the metaphorical analysis route has also been mooted (see, e.g., [32]), and would gain weight in a richer case such as "The country is sweating with the effort of getting rid of slavery."

[8] (3) apperas to assume that Earth's physical moon has a dark half that cannot be seen. here there seems to be a mistaken supposition that the dark side is a fixed part of the moon, rather than changing as the moon orbits the Earth. Also, the passage may be mistakenly equating the dark side with the side facing away from the earth.

is no warrant for taking that side to be bright and salient. But, the fact that the message cannot come just from the second clause alone is a not a reason for saying that the first clause should be given its own metaphorical meaning, but is rather a reason to say that a unified fictional scenario should be constructed from both clauses, and then target-scenario meaning should be extracted from that scenario as appropriate. However, I do not have a specific theory about how hearers are pressured to adopt this more holistic approach across clauses/sentences and when they give them separate metaphorical meanings.

Thirdly, I have sought to explain chained metaphor (where something A is viewed as B and something about B is viewed as C) in terms of nesting of fictions within each other. I have treated some real examples elsewhere, but a simple, chained variant of (1) would be "The managers had cricks chewing into their necks ...." where the managers' state is metaphorically cast as having a crick in their necks but the cricks are in turn cast as being animals. This would be handled by having the fictional scenario discussed above, but now there would be, nested within it, a fiction in which the cricks are animals. This nesting is of course similar to the common phenomenon of stories-within-stories. It would appear that this matter needs further attention in the philosophy of fiction (not least because of the question of whether or not it is merely fictional that the inner fiction exists, and how one formally cashes out that potential meta-fictionality), while on the other hand metaphor research has been slow to come up with detailed theories of chained metaphor.

3 AN ANTI-ANALOGY-EXTENSION THESIS

In the ATT-Meta approach, as in conceptual metaphor theory, metaphor is based on familiar analogies. An ATT-Meta metaphorical view involves a set of entrenched analogical ncorrespondenceXX rules, and VNMAs are additional analogical correspondence rules. Nevertheless, a key point about the ATT-Meta approach can be called the *Anti-Analogy-Extension Thesis*.[9] This says that open-ended view-transcending elements of the source subject matter (e.g., the crick in (1), the insisting in (2)) should *not*, normally, be given target-scenario parallels, and in particular that existing analogies should not be extended to encompass those elements—they should be left unparalleled. ATT-Meta seeks to get away with the least amount of analogy possible, *contra* other theories such as Structure-Mapping Theory [22, 15], which assume that the task is to maximize the extent of analogy.

In contrast to such theories, the ATT-Meta approach claims that the hearer tries to connect view-transcending to within-fiction content that *can* be converted via already-known correspondences (view specific or view-neutral). This is on the theoretical principle that, typically, the unparalleled items are proposed by a speaker not as individually standing for aspects of the target scenario being addressed, but rather to build a fictional scenario that holistically illuminates the target side using correspondences that the hearer is expected already to know.

In particular, in the neck-crick example (1), the cricks and resultant physical pain have no parallel in the target scenario. The cricks are only there to convey emotional distress, difficulty in continuing the conversations, etc. Similarly, there is no need at all to propose that for (2) the mentioned part corresponds to an identifiable aspect of the real person, or to propose that there is some internal, real mental action that can be clearly held to correspond to the action of insisting in the sentence. Rather, the mentions of a part and of insisting

---

9 The account in this section is based on [8].

are *merely* tools towards constructing a rich fictional scenario, which in turn conveys in an economical, accessible and vivid manner the possession of a particular sort of mental state by Mary.

The Anti-Analogy-Extension Thesis goes hand in hand with a form of holism about the fictional scenarios and the metaphorical sentences leading to them, related to the holism of the previous subsection. The fictional scenario is to be regarded not as having a detailed analogy to a target scenario but rather something that *holistically* conveys information about the target scenario. This conveying is, to be sure, done by the action of correspondences that pick on specific aspects of the fictional scenario. But the ultimate intent here is to transfer information, not specify an analogy. And any specific aspect of the fictional scenario that is grabbed by a correspondence may be the result of inference over large amounts of information within the scenario. In particular what this means is that there may be no specific part of the metaphorical sentences that can be said to correspond to a given aspect of the reality scenario (although this can happen in simple cases of metaphor). For example, going back to (2), an aspect of its meaning not detailed above (but explained in [9]) is that Mary lacks the belief that Mike is adorable (she merely has a tendency to believe it, and indeed also has a tendency to disbelieve it). This lack does not correspond to any one aspect of (2) but rather to the whole of (2).

Another work that emphasizes both frequent holism of metaphor (in this subsection's sense) and the lack of need for, or indeed the frequent undesirability of, analogy-extension is Langlotz's treatment of idioms [30], including metaphor-based ones.

4 METAPHORICITY OF SOME THOUGHT

The anti-analogy extension thesis has interesting consequences for the nature of thought, consequences that have barely been addressed in AI or philosophy and need more work in metaphor theory itself. Within the cognitive linguistics field, it is typical to think of metaphor as something that is somehow fundamental in the mind, not just in communication and external expression, and in particular to think of many concepts, particularly abstract ones, as in some way structured by metaphor (i.e., by being linked by metaphorical mappings to source concepts). See [40] and [33] for critical discussion of some of the main points here. One reason for the hypothesis is that metaphor occurs in media other than language, such as in graphical media. One might try to account for this in a number of ways, but an one parsimonious option is that metaphor is inherently a mental as opposed to purely communicative or externally-expressive phenomenon. I will take the point to basically be that, when thinking but not externally communicating about some subject matters, we are at least sometimes mentally using metaphorical mappings between those subject matters and suitably-related source subject matters. There is no implication here that this mental activity is conscious. I assume here that it may well be unconscious.

The Anti-Analogy-Extension Thesis leads to an especially strong claim: namely, that major portions of a metaphorical thinking episode may not individually have *any* translation into non-metaphorical thoughts within the person's mind. This is because extensive areas within a metaphorical fiction may not have any analogical correspondence to the target scenario, but rather just serve indirectly to support those limited aspects of the fiction that are in analogical correspondence to the target. Open-ended elaboration of fictional scenarios could exist in mind just as much (or more) than in language and other external expression. For example, someone thinking (but not communicating) about the managers in (1) may

mentally develop the fictional scenario in creative ways as above, such as imagining pains in many parts of the managers' bodies, not just their necks, imagining the managers massaging those parts, contorting themselves, etc. These could have consequences about the intensity of the emotional states, their longevity and difficulty of eradication, and the desires of the managers. These conclusions can be mapped to reality. But most of the fictional scenario is *not* mapped.

I also wish to make a more radical conjecture. In the discussion so far, even if some thoughts are in an unparalleled region of a fictional scenario, their function in the mind is nevertheless to support fiction-to-target conversions that produce mental representations directly in terms of the target subject matter. One might say that the latter representations are literally about the target scenario—so the unparalleled parts of the fiction are indirectly connected to those literal representations. But it is possible that there are metaphorical representations in the mind that have *no* connection to a literal description of the target scenario, even indirectly. For instance, one can conceive of a person whose only resource for thinking about electricity is that it is a liquid flowing within wires, etc. She knows nothing about electricity other than what can be approximately captured by these resources, and she has no translation of the liquid-based thoughts about electricity into any other terms. Many of our concepts about relatively abstract matters, such as time, electricity, money, love, mental states, ... at least *include* metaphorical views, and I am now supposing that a concept could consist *only* of such a view. So, the person's concept is *irreducibly metaphorical*. (This does not mean either that it is irreducible in principle or that for some other person it is not irreducible.)

Yet the person might agree, if asked, that electricity isn't *really* a liquid. If she knows about metaphor, she might more specifically agree that electricity is only metaphorically a liquid. So, we as observers, and even the person herself, should not take her to think that electricity really is a liquid, but rather as metaphorically thinking about electricity as a liquid, perhaps unconsciously. As long as her liquid thoughts are adequately linked to relevant actions she needs to take in the world (e.g. actions on switches, carefulness about cutting wires, etc.) she can operate in the world perfectly well for everyday purposes. While this sort of possibility falls naturally out of standard cognitive linguistic considerations (even if not yet fully developed in that field), it appears not to be catered for in detailed theories of representation and mind in AI and philosophy.

5 ATTACKING AN ESOTERIC NETTLE WITH THE SCYTHE OF METAPHOR

I believe considerations of metaphor can help with a long-standing philosophical problem about the nature of propositional attitudes (broadly, contentful thoughts) and the meaning of propositional attitude reports—reports of mental states, with sentences of the form "John believes ..." as the simplest sort of example. Metaphor could provide a radically new, and subversive, solution. I call the problem one of *esoteric imputation*. It has been noted in different forms by various philosophers, such as Clapp, Richard, Schiffer and Soames (see citations below), and often arises with attempts to provide theories of propositional attitudes (PAs) and the meaning of PA reports. The problem is that theories are in danger of imputing, to ordinary people, thoughts that implausibly involve esoteric aspects of non-commonsensical explications of thought that are postulated by the theories.[10]

For example, one common type of theory is roughly that the meaning of "John believes that spies are evil" is that John is in a certain

relation BEL to the proposition that spies are evil, via some "mode of presentation," "way of thinking" or "guise" for that proposition. Such a theory involves some specific, technical notions of matters such as what a proposition is, what a mode of presentation (etc.) is, what it is for a mode of presentation to present something, what BEL is, and what it is for a proposition to refer to the world. Typically, while some aspects of these technical notions might be reasonably intuitive, the whole package is so esoteric that it is unimaginable that anyone other than philosopher could entertain them in their thoughts.[11] (See [36, 37] for complaints along these lines, in discussion of the "meaning intention problem." See also [1].) Lest someone think that what one calls the meaning of a PA report or any other sentence needn't be the same thing as the content that a hearer grasps when encountering it, I should point out that the problem arises also in iterated attitude reports such as "Mary believes that John believes that spies are evil." Here, one's theory of PAs and PA reports should not have as a consequence that Mary has a belief that is couched in terms of of the esoteric explication of John's belief that the theory would assign as the meaning of "John believes that spies are evil." or more broadly as the scientifically accurate nature of what it is for John to believe that spies are evil.

Some specific further instances of the problem arising in the philosophical literature are as follows, interlaced with some observations of my own. Schiffer [37, pp.35–37] highlights an esoteric imputation problem with Fregean accounts of PA reports, in that belief reporters are unaware of the detailed nature of concepts, and notably of Fregean ones. Hornstein [27] characterizes many PA theories as requiring the belief reporter to have some grasp of theories of sense and reference, and he implies that this is mysterious. Edelberg [20] says that an approach by Kaplan to PA reports seems implausibly to require ordinary people to know and understand Kaplan's theory. Braun [16] suggests that the hypothesized speaker thoughts about modes of presentation in the above approach cannot be made explicit by speakers, casting doubt on the existence of those thoughts. Berg [14, pp.26–27] worries that an explanation of what it is to believe a proposition *under* a given mode of presentation is (what I would call) esoteric. Clapp [19] makes claims about major PA report accounts requiring speakers to know esoteric things about ordinary believers' thoughts, and he claims that attempts to mitigate this problem don't fully work and/or make the accounts fall into other problems. Clapp implies that even the authors who are aware of such [esoteric imputation] problems have failed to solve the problem.

To get some of the flavour of current discussion about the topic, we can consider Richard's [35, Ch.13] response to a complaint by Soames [38, p.170] against his account. Soames questions whether speakers really intend to commit themselves to complex claims (that he takes Richard's theory to involve) about the languages or internal mental representations used by believers to which they typically ascribe beliefs. Richard counters that the thoughts he is imputing to speakers are in fact not implausibly complex; and I also take him to argue that the thoughts are not esoteric. He says "it is uncontroversial that conversants routinely make presuppositions about how others represent the world[.]" This may be true but the question really is whether conversants have the particular sorts of thoughts about the particular sorts of representations that Richard proposes. I am made

---

[10] This section draws from [10].

[11] At least, it's unimaginable that they can consciously do so, and only with a theory that radically dislocates unconscious from conscious thought would allow them to unconsciously think in terms of such esoteric notions even though they cannot do so consciously. (My impression is that the tension here between unconscious and conscious thought is not commonly enough considered in the philosophical area in question.)

nervous by the following statement by Richard [35, Introduction, p.22], concerning a report of form "Boswell thinks that S." According to Richard's theory, this has a logical form that can be glossed in English along the following lines, where "annotated proposition" is a technical, rather esoteric notion that Richard has defined:

> There's an acceptable translation manual ... such that one of Boswell's beliefs (i.e. an annotated proposition determined by one of his belief states) is translated, under that translation manual, by the annotated proposition that S.

So, suppose we consider Yolanda believing that Boswell thinks that S. Does she then have something like the concept of a mental translation manual or of an annotated proposition? Perhaps it is plausible that she has such thoughts, via suitable modes of presentation perhaps, but it is up to Richard to convince us of it.

Also, the book by King, Soames and Speaks [28] contains several comments relevant to esoteric imputation. For instance, Soames's and Speaks's articles in the book complain that King's account there requires ordinary language users to have esoteric thoughts. But Soames's account in the book has, itself, an esoteric imputation problem. It is central to his proposal that people become familiar with their own cognitive acts and then abstract from these to become familiar with more general, agent-independent cognitive-act types (constituting propositions etc.). But I suspect that individual act types as portrayed by Soames are esoteric: certainly, discussions in the literature about them are highly esoteric. Also, if people's categories are generally based on prototypes and/or exemplars, then this may apply just as much to cognitive-act types as to other types of things; but then it becomes difficult to isolate objectively existing act types of Soames's sort.

Thus, we have evidence that it is extremely difficult to come up with theories that avoid esoteric imputation problems using current philosophical resources. While it may yet be possible to do so, it would appear to involve theoretical contortions of great agility and knottedness. In response, I suggest a different strategy, inspired by the claim in cognitive linguistics and elsewhere that people often conceive of mental states, along with many other abstract matters, with the help of metaphor. I suggest that PA theory should positively impute to ordinary agents thoughts about each other's mental states and processes that are *framed in terms of commonsensical metaphor*. The basic idea is that a hearer of, say, "John believes that spies are evil" will (typically unconsciously) think of John's mental state in a metaphorical way, e.g. by thinking of John saying something to himself (silently) in English, or as John having having a mental image of spies being evil, or some combination of these. Equally, in an iterated case such as "Mary thinks that John believes that spies are evil," the hearer imputes to Mary a metaphorical view of John's mental state. Of course, there is an important question here about what particular view or views Mary might impute to John. I discuss this in [10].

In short, the advocated approach *deliberately* imputes to ordinary people *commonsensical, metaphorical* thoughts about mental states, rather than *non-deliberately* imputing to them *non-commonsensical, esoteric* thoughts about mental states. Particular effects of this approach. apart from avoidance of esoteric imputation, include (a) a new range of ways in which believing (or hoping, wanting, ...) in general may be viewed in acts of attitude report understanding, and (b) metaphor-relativity in the distinctions between different styles of interpretation such as transparent and opaque, which have been much discussed in the philosophical and AI literatures as if they were objectively characterizable.

Naturally also, insofar as the metaphorical framing of a situation affects one's behaviour in/towards it, the approach has practical consequences for AI systems that are meant to be interacting with human beings who are having thoughts about other people's thoughts.

6 COGNITIVE ADDITION OF METAPHOR IN LANGUAGE UNDERSTANDING

The approach to propositional attitude reports advocated in the previous section rests on an assumption that metaphor can be *cognitively added* during understanding. The hearer's understanding of the sentence "John believes ..." is *metaphorically* couched in the hearer's mind, even though the sentence itself contains nothing that would typically be called metaphorical by metaphor researchers. Thus, metaphoricity has been added by the hearer. But this isn't a special assumption just to make that approach work. It arises very naturally out of much more general considerations.

Recall the view in cognitive linguistics that metaphor is a conceptual matter, not primarily a matter of language or other modes of external expression. For instance, it is supposed that people think about time using any of a variety of metaphorical views (see, e.g., [34]). Under one, the person is moving along a spatial axis towards events, and in a dual of this, events are moving toward the person. There has been much discussion of the use of such views in interpreting metaphorical sentences such as "The meeting was moved forward/back." However, my claim is that the interpretation even of a *literal* sentence such as "The meeting time was changed to noon on the next day" can be accompanied by metaphorical couching of what the sentence says. If the hearer's concept of and general private thoughts about time include metaphorical aspects (even if not irreducibly so) it is only natural to suppose that those aspects are activated even by literal utterances about time. Thus, for the sentence "The meeting time was changed to noon on the next day" the hearer may mentally construct a metaphorically couched thought that paints the meeting as having been moved along a spatial axis.

Recent work in empirical psycholinguistics such as in [23, 24] suggests that people do often activate concepts in the source domain of a metaphorical view when understanding a *metaphorical* utterance based on it. This can even happen when the metaphorical language is highly conventional or even supposedly "dead." It is not a big step from here to the idea that people also do cognitive addition of metaphor when understanding some literal language (which is often "dead" metaphor anyway).

But it appears that all work on metaphor within language in philosophy and AI is confined to the question of how to account for the meaning of sentences that are, so to speak, already metaphorical. There appears to be an uncritically adopted, tacit assumption that the understanding of an ostensibly literal sentence only ever involves semantic representations that are themselves directly about the subject matter at hand, rather than bearing a metaphorical or other indirect relationship to that subject matter. But in reality we must countenance the possibility that the figurativeness or otherwise of utterances is only weakly related to the figurativeness or otherwise of the mental representations arising from or giving rise to the utterances.

7 CONCLUSION

I commend the issues covered in this paper as possible discussion points for Computing & Philosophy researchers who are interested in metaphor or foundational issues concerning the meaning of language.

The different sections above depend on each other to a considerable extent, although there are islands of independence. The anti-analogy-extension thesis is facilitated by a fiction-based account, and perhaps requires such an account. Thus the particular points made about metaphor within thought, which exploit that thesis, also depend on a fiction-based approach (but other approaches could also embrace metaphor in thought in other ways). However, the general notion of cognitive addition of metaphor does not presuppose a fiction-based approach. The use of metaphor to address the esoteric imputation problem for propositional attitude theory assumes that thought can be metaphorical and that cognitive addition happens. In fact it assumes, though this was not explicitly stated above, that a person's X's thoughts about other people's thoughts are often irreducibly metaphorical, and this does amount to viewing X's thoughts as defining fictions that are not cashed out in non-fictional target scenarios in X's mind.

ACKNOWLEDGEMENTS

REFERENCES

[1] Bach, K. (2000). Do belief reports report beliefs? In K.M. Jaszczolt (Ed.), *The Pragmatics of Propositional Attitude Reports,* pp.111–136. Oxford: Elsevier.

[2] Barnden, J.A. (2001a). Uncertainty and conflict handling in the ATT-Meta context-based system for metaphorical reasoning. In V. Akman, P. Bouquet, R. Thomason & R.A. Young (Eds), *Modeling and Using Context: Third International and Interdisciplinary Conference* (CONTEXT 2001). Lecture Notes in Artificial Intelligence, Vol. 2116, pp.15–29. Berlin: Springer.

[3] Barnden, J.A. (2001b). Application of the ATT-Meta metaphor-understanding approach to various examples in the ATT-Meta project databank. Technical Report CSRP–01–02, School of Computer Science, The University of Birmingham, U.K.

[4] Barnden, J.A. (2006). Consequences for language learning of an AI approach to metaphor. In J. Salazar, M. Amengual & M. Juan (Eds), *Usos Sociales del Lenguaje y Aspectos Psicolingüísticos: Perspectivas Aplicadas,* pp.15–57. Palma, Mallorca: Universitat de les Illes Baleares.

[5] Barnden, J.A. (2008). Metaphor and artificial intelligence: Why they matter to each other. In R.W. Gibbs, Jr. (Ed.), *The Cambridge Handbook of Metaphor and Thought*, 311–338. Cambridge, U.K.: Cambridge University Press.

[6] Barnden, J.A. (2009). Metaphor and context: A perspective from artificial intelligence. In A. Musolff & J. Zinken (Eds), *Metaphor and Discourse*, pp.79–94. Basingstoke, UK: Palgrave Macmillan.

[7] Barnden, J.A. (2010). Metaphor and metonymy: Making their connections more slippery. *Cognitive Linguistics, 21*(1), pp.1–34.

[8] Barnden, J.A. (2015). Open-ended elaborations in creative metaphor. Invited chapter for Besold, T.R., Schorlemmer, M. & Smaill, A. (Eds,) *Computational Creativity Research: Towards Creative Machines,* pp.217–242. Atlantis Press (Springer). (Series: Atlantis Thinking Machines, Vol. 7.)

[9] Barnden, J.A. (in press). Mixed metaphor: Its depth, its breadth, and a pretence-based approach. In R.W. Gibbs, Jr. (Ed.), *Mixed Metaphor*. Amsterdam: John Benjamins.

[10] Barnden, J.A. (under review). Propositional attitudes and commonsense psychology: Using metaphor to slay an esoteric nettle. Submitted to a philosophy journal.

[11] Barnden, J.A., Glasbey, S.R., Lee, M.G. & Wallington, A.M. (2004). Varieties and directions of inter-domain influence in metaphor. *Metaphor and Symbol, 19*(1), pp.1–30.

[12] Barnden, J.A. & Lee, M.G. (2002). An artificial intelligence approach to metaphor understanding. In Tomasz Komendzinski (Ed.), "Metaphor: A Multidisciplinary Approach," a special issue, *Theoria et Historia Scientiarum, 6* (1), pp.399–412.

[13] Barnden, J.A. & Wallington, A.M. (2010). Metaphor and its unparalleled meaning and truth. In A. Burkhardt & B. Nerlich (Eds), *Tropical Truth(s): The Epistemology of Metaphor and Other Tropes,* pp.85–121. Berlin / New York: De Gruyter.

[14] Berg, J. (2012). *Direct belief: An essay on the semantics, pragmatics and metaphysics of belief.* Berlin/Boston: De Gruyter Mouton.

[15] Bowdle, B.F. & Gentner, D. (2005). The career of metaphor. *Psychological Review, 112*(1), pp.193–216.

[16] Braun, D. (1998). Understanding belief reports. *The Philosophical Review, 107* (4), pp. 555–595.

[17] Brians, P. (2003). *Common errors in English usage.* Franklin, Beedle & Associates, Inc.

[18] Carston, R. & Wearing, C. (2011). Metaphor, hyperbole and simile: A pragmatic approach. *Language and Cognition, 3*(2): pp.283–312.

[19] Clapp, L. (2000). Beyond sense and reference: An alternative response to the problem of opacity. In K.M. Jaszczolt (Ed.), *The Pragmatics of Propositional Attitude Reports,* pp.43–75. Oxford: Elsevier.

[20] Edelberg, W. (1992). Intentional identity and the attitudes. *Linguistics and Philosophy, 15*, pp.561–596.

[21] Fauconnier, G. & Turner, M. (2008). Rethinking metaphor. In R.W. Gibbs, Jr. (Ed.), *The Cambridge Handbook of Metaphor and Thought*, pp.53–66. Cambridge, U.K.: Cambridge University Press.

[22] Gentner, D. (1983). Structure-mapping: a theoretical framework for analogy. *Cognitive Science, 7* (2), 95–119.

[23] Gibbs, R.W., Jr & Matlock, T. (2008). Metaphor, imagination, and simulation: Psycholinguistic evidence. In R.W. Gibbs, Jr. (Ed.), *The Cambridge Handbook of Metaphor and Thought*, pp.161–176. Cambridge, U.K.: Cambridge University Press.

[24] Gibbs, R.W., Jr & Santa Cruz, M.J. (2012). Temporal unfolding of conceptual metaphoric experience. *Metaphor and Symbol, 27*(4), pp.299–311.

[25] Goatly, A. (1997). *The language of metaphors.* London and New York: Routledge.

[26] Grady, J.E. (1997). THEORIES ARE BUILDINGS revisited. *Cognitive Linguistics, 8*(4), pp.267–290.

[27] Hornstein, N. (1984). *Logic as grammar.* Cambridge, MA: MIT Press.

[28] King, J.C., Soames, S. & Speaks, J. (2014). *New thinking about propositions.* Oxford: Oxofrd University Press.

[29] Lakoff, G. & Johnson, M. (2003). *Metaphors we live by.* 2nd Ed. Chicago: University of Chicago Press.

[30] Langlotz, A. (2006). *Idiom creativity: A cognitive-linguistic model of idiom-representation and idiom-variation in English.* Amsterdam/Philadelphia: John Benjamins.

[31] Levin, S.R. (1988). *Metaphoric worlds.* New Haven, CT and London, U.K.: Yale University Press.

[32] Low, G. (1999).“This paper thinks...": Investigating the acceptability of the metaphor AN ESSAY IS A PERSON. In L. Cameron & G. Low (Eds), *Researching and Applying Metaphor,* pp.221–248. Cambridge, U.K.: Cambridge University Press.

[33] Murphy, G.L. (1996). On metaphoric representation. *Cognition, 60*, pp.173–204.

[34] Moore, K.E. (2006). Space-to-time mappings and temporal concepts. *Cognitive Linguistics, 17*(2), pp.199–244.

[35] Richard, M. (2013). *Context and the attitudes: Meaning in Context, Vol. 1.* Oxford: Oxford University Press.

[36] Schiffer, S. (1992). Belief ascription. *J. Philosophy, 89* (10), pp.499–521.

[37] Schiffer, S. (2003). *The things we mean.* Oxford: Oxford University Press.

[38] Soames, S. (2002). *Beyond rigidity.* Oxford: Oxford University Press.

[39] Stern, J. (2000). *Metaphor in context.* Cambridge, MA and London, UK: Bradford Books, MIT Press.

[40] Vervaeke, J. & Kennedy, J.M. (2004). Conceptual metaphor and abstract thought. *Metaphor and Symbol, 19*(3), pp.213–231.

[41] Waggoner, J.E. (1990). Interaction theories of metaphor: psychological perspectives. *Metaphor and Symbolic Activity, 5* (2), pp.91–108.

[42] Walton K. (2004/1990). Fiction and non-fiction. In E. John & D.M. Lopes (Eds), *Philosophy of Literature—Contemporary and Classic Readings: An Anthology*, pp.136–143. Oxford: Blackwell. Reprinted from *Mimesis and Make-Believe: On the Foundations of the Representational Arts,* pp.36–41, 70–73, 77–78, 81–82, 84–89, Cam., Mass: Harvard UP, 1990.