AISB Convention 2015

University of Kent, Canterbury

Proceedings of the AISB 2015 Symposium on Computing and Philosophy: The Significance of Metaphor and Other Figurative Modes of Expression and Thought

Edited by John Barnden and Andrew Gargett

Introduction to the Convention

The AISB Convention 2015—the latest in a series of events that have been happening since 1964—was held at the University of Kent, Canterbury, UK in April 2015. Over 120 delegates attended and enjoyed three days of interesting talks and discussions covering a wide range of topics across artificial intelligence and the simulation of behaviour. This proceedings volume contains the papers from the *8th AISB Symposium on Computing and Philosophy: The Significance of Metaphor and Other Figurative Modes of Expression and Thought*, one of eight symposia held as part of the conference. Many thanks to the convention organisers, the AISB committee, convention delegates, and the many Kent staff and students whose hard work went into making this event a success.

—Colin Johnson, Convention Chair

Introduction to the Symposium

Communication and expression in language, pictures, diagrams, gesture, music etc. is rich with figurative aspects, such as metaphor, metonymy, hyperbole and irony. People engage in such communication and expression in a variety of contexts and with a range of effects. Modelling figurative patterns of communiction/expression is a key aim of academic disciplines such as linguistics, philosophy, discourse studies, and psycholinguistics, and automatically understanding such phenomena is a long-standing and now expanding endeavour within Artificial Intelligence, with metaphor generation also receiving more attention. In addition, some researchers have suggested that metaphor can be an intrinsic part of thought, not just an aspect of external communication/expression.

Specific topics of interest for the Symposium included the following:

- How philosophical thinking on figurative expression and thought can/should be exploited/heeded by relevant AI researchers
- How computational attempts to model figurative expression can aid philosophical thinking about it
- How the production of figurative expression reflects speakers' conceptualisations, goals and commitments
- How to model/analyse/understand the emotional and evaluative content of figurative expression
- The intersection of issues of figurative expression and issues of embodiment, enactivism, cognitive simulation, etc.
- Whether thought, as opposed to external expression, can be metaphorical, ironic, etc., and if so what this amounts to (philosophically, computationally, psychologically, ...)
- How figurative and especially metaphorical thinking might be involved in introspection, and therefore be bound up with the nature of consciousness
- Links between figurative thought/expression and the nature of creativity
- Figurative aspects of philosophical theorizing (about any topic), especially as uncovered by detailed technical analysis of figuration
- Figurative aspects of notions of computation…and even: could the notion of computation be irreducibly metaphorical?

It was a specific aim of the Symposium to encourage speculative thought, provisional proposals, and provocative question-raising based on careful analysis of issues. The papers in this volume serve that aim well. Given the broad scope of the Symposium we could not hope to cover all the topics listed above, but the papers range widely and bravely over the remit of the Symposium, which brings together disciplines in an unusual way.

We thank Colin Johnson as Convention chair, and the Committee of the AISB, for providing the opportunity to hold a symposium with the above remit. We thank our co-organizers Mark Bishop and Yasemin Erden for their help, and moreover for entrusting us with an event in a symposium series that they have played leading roles in over the years. We also thank our Programme Committee, which had wide international and disciplinary reach, for their hard work. Apart from the organizers, the committee contained Tony Beavers, Jerry Feldman, Eugen Fischer, Mark Phelan, Mihaela Popa, Mark Sprevak, Tony Veale and Yorick Wilks.

—John Barnden and Andrew Gargett, Symposium Organisers

Contents

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

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.

Metaphor and understanding *me*

Yasemin J. Erden [1]

Abstract. This paper explores the role of the metaphor-maker in the construction of meaningful metaphor construction. More specifically, the paper defends the claim that the semantic-language-user is key for the possibility of both meaning and the understanding of metaphor. This takes into account the seemingly contradictory status of two claims: (1) that words can be meaningful without context, intentionality or the presence of, or origin in a language-user, while (2) the expectation of a context, intention or speaker is central to finding meaning in words and particularly metaphors. The apparent contradiction can be resolved if we see that the possibility of meaningful metaphor says as much about our expectation and need for meaning as it does about the language itself. Understanding words is thus as much about understanding the utterer of the words, as about the words themselves. Through exploring Wittgenstein's ideas about metaphor, this idea should become clearer. The paper will then explore what the limitations of computational metaphor might be as a result.

1 INTRODUCTION

What does it mean to understand a person through their words? And what do words mean separate from a speaker? These are questions that this paper explores in order to understand the central question: how are metaphors *meaningful?* In this, the aim is not to discuss the meanings of individual words, but rather to explore the very possibility of meaning and to point to the central roles played by context, expectation, experience and embodiment. To do this we begin by looking at a short quotation from Ludwig Wittgenstein, which has puzzled commentators because of its self-referential turn of phrase. The claim is made (or rather, defended, since the claim is not new even if it remains controversial) that to understand the phrase requires that we understand the person, Wittgenstein, as well as the words he uttered in that sentence [2].

Building on this, I argue that the possibility of a meaningful metaphor relies on context within which language is embedded, such as described by Wittgenstein [3] in terms of *language-games*. This does not lead to a strong claim that computational-metaphor is impossible though it does suggest a weaker claim that to be *successful* (which includes indicators such as 'appropriateness' or even 'acceptable') in this area may be tricky. This is partly because what is considered either appropriate or acceptable in ordinary language is already tricky (including where highly creative language-use can muddy the waters of ordinary language substantially). It is also partly because of the role that *expectation of meaning* creates. As I discuss elsewhere [4] [5], meaningful language-games require not only a successful meeting of *rules*, but also a willing on the part of participants to *recognise* other speakers as meaningful language-users. In the case of the words uttered by Wittgenstein, it is precisely because scholars expect meaning to be found, that the search for a meaning is considered worthwhile.

To explore this further, we will also discuss the possibility of non-human (or computational) metaphor construction,

interpretation and use, and discuss the likely limitations that may occur where such construction is disembodied and decontextualized. The concept of the language-game will be employed in this discussion, since Wittgenstein offers this as a metaphor for meaningful language use. The metaphor of a *game* is particularly helpful for exploring ideas about participation and mimicry, and thereby how we view the relationship between computational and non-computational approaches to both metaphor understanding and production. Will we accept a metaphor as creative or even useful if we do not believe the person (or program) has any idea (understanding or experience) of the individual components, let alone the comparison being drawn?

Finally, discussion will explore the way that, on the one hand we might measure the *success* of a program (in constructing or interpreting metaphorical language) according to a set of pre-determined rules (even if these can be later amended or more fully altered), while on the other hand, the idea that we can accept or reject metaphors based on issues aside from content, including context and expectation of meaning. An unusual or bizarre comparison might make sense where we look for (or expect) sense, for example from a person who I know uses and *understands* the same language as me, and not where we expect little sense to be found, such as in the babbling of a small infant. The expectation of meaning is an important element in drawing these sorts of comparisons, and can sometimes be unfair in the expectation (or not) of meaning and importantly for this discussion, in what is then accepted as either meaningful or indeed successful.

1 UNDERSTANDING *ME*

In proposition 6.54 of the *Tractatus Logico-Philosophicus* [1] (first published in 1921), Ludwig Wittgenstein states of his project: "My propositions serve as elucidations in the following way: anyone who understands me eventually recognises them as nonsensical." Understanding what Wittgenstein meant by these simple yet enigmatic words has dominated certain sub-sections of Wittgenstein scholarship. In one particular strand of scholarship, discussion centres on that little word "me" and why Wittgenstein did not instead write, "understands *them*" in reference the propositions of the text, as per the second half of his statement. Understanding why this might be important will have an impact on the arguments of this paper.

This paper picks up this discussion in order (in the first instance) to lend support to the interpretation offered by Cora Diamond [2, p. 151] whereby to understand this statement requires that we understand both Wittgenstein as well as his words. She claims this is a clear indication that Wittgenstein wanted to "draw attention to a contrast between understanding a person and understanding what the person says." This, she says, is pivotal for our understanding of the instruction that Wittgenstein presents in these words, which is that we should recognise the propositions of his text as nonsensical. This seeming contradiction puzzles, delights and infuriates readers

often in equal measure. How can the propositions be taken as nonsense if we can in fact understand them?[1] In following Diamond's solution we dissolve the contradiction since we can accept (if we like) that the *content* of the *Tractatus* is nonsense, while simultaneously acknowledging that we have somehow understood this nonsense because we understand the person. Thus we come to 'understand not the propositions but the author' [2, p.155].

One objection to this view, such as is offered by Priest [6, p. 150], argues that the conclusion of the nonsense uttered (and so-called) in the Tractatus results only in a contradiction. Regardless of context, it is clear that we have at some point understood nonsense—it must have made sense to us—otherwise what did we understand? Yet Diamond's reply to such arguments is that although we have seemingly understood what is later termed *nonsense*—Priest is not wrong in this—this does not mean it is any the less nonsensical. In fact, she holds [2, p. 150] it is not that we understood the nonsense propositions in the first instance, thus generating a contradiction, but rather that "in recognising that they are nonsense, [we] are giving up the idea that there is such a thing as understanding them". She concludes, "What Wittgenstein means by calling his propositions nonsense is not that they do not fit into some official category of his of intelligible propositions but that there is at most the illusion of understanding them". The reason for this approach, she claims, hinges on seeing Wittgenstein's request that we understand *him* as indicative of his personal engagement with those who talk nonsense, something she later describes [2, pp. 157-58] as requiring imagination:

> My point then is that the Tractatus, in its understanding of itself as addressed to those who are in the grip of philosophical nonsense, and in its understanding of the kind of demands it makes on its readers, supposes a kind of imaginative activity, an exercise of the capacity to enter into the taking of nonsense for sense, of the capacity to share imaginatively the inclination to think that one is thinking something in it. If I could not as it were see your nonsense as sense, imaginatively let myself feel its attractiveness, I could not understand you. And that is a very particular use of imagination.

This recourse to imagination is perhaps surprising (and is not itself uncontroversial or indisputable), but it is helpful for when we consider ideas about analogy, and more specifically metaphor, to which we now turn.

2 AN EXPECTATION OF MEANING

The discussion above offers a way in which to begin to see that the possibility of meaningful language and understanding relies on such words having been uttered by a semantic language-user (in the above example, Wittgenstein). In fact, the crux of this paper, where metaphor is concerned, is that people (lay- and scholars alike) would not have been so interested in the enigmatic aphorism noted above if the speaker had not been a person. If Wittgenstein had instead been the name of a complex computational program that uttered such words, it is unlikely the discussion about them would have lasted nearly a hundred years. More simply: if Wittgenstein had been a machine, we'd likely

have ignored the odd turn of phrase, or perhaps described as a superficial error.

This approach to understanding an author over (or at least as well as) her/his words may seem in contrast to Barthes [7] and related post-structuralist ideas about the independence of text from an author (commonly referred to, in reference to Barthes, as *the death of the author*). However, the *death* of an author does not thereby presume *no* author. Instead the argument is a complex response to some traditional notion of the *individual*— the author—as the final locus of meaning. In other words, the authorial voice as judge, authority, "always finally the voice of one and the same person, the author, which delivered his 'confidence'" [7]. As he notes elsewhere, the crux is to do with culture, which is akin to context that I describe above:

> We know that a text does not consist of a line of words, releasing a single "theological" meaning (the "message" of the Author-God), but is a space of many dimensions, in which are wedded and contested various kinds of writing, no one of which is original: the text is a tissue of citations, resulting from the thousand sources of culture. [7]

The text and the author exist *simultaneously* on this account, and in this way, the text has as much authority as the author, the reader, and any other voice in dialogue about the text. "In this way is revealed the whole being of writing: a text consists of multiple writings, issuing from several cultures and entering into dialogue with each other" [7].

While this would seem to stand in tension to the discussion about Wittgenstein's text above—where we should understand Wittgenstein in order to understand the text—in fact we can see the same impetus of the centrality of the reader's voice in Wittgenstein's work also. In the Preface to the *Tractatus*, Wittgenstein says, "This book will perhaps only be understood by those who have themselves already thought the thoughts which are expressed in it—or similar thoughts. It is therefore not a text-book. Its object would be attained if it afforded pleasure to one who read it with understanding" [1]. Furthermore, the claim to a singular authorial voice is never made. As he explains a little further along, "How far my efforts agree with those of other philosophers I will not decide. Indeed what I have here written makes no claim to novelty in points of detail; and therefore I give no sources, because it is indifferent to me whether what I have thought has already been thought before my by another" [1]. Similar to Barthes, the authorial voice is not to be considered that of an individual in any absolute sense, or a decontextualised authority. Instead we can take Wittgenstein's words, his contribution to the dialogue, as direct engagement with, and an imploring to, the reader to understand. His request at the end of the text that we understand *him* specifically, is as much a part of this collective, contextual engagement, as Barthes' claims that,

> the unity of a text is not in its origin, it is in its destination; but this destination can no longer be personal: the reader is a man without history, without biography, without psychology; he is only that someone who holds gathered into a single field all the paths of which the text is constituted. [7]

This is not to say that there are no differences between their respective views however, and indeed I will return to this in Section 3 below.

From this we arrive back at the discussion above regarding context, and to this we can add shared experience, culture, history and meaning. For these reasons I offer the claim that the

---

[1] It is important to clarify that the author does not in fact take at face value the nonsensicality of the propositions in Wittgenstein's text, but this argument is outside the scope of this paper.

AISB Convention 2015: 8th AISB Symposium on Computing and Philosophy: The Significance of Metaphor and Other Figurative Modes of Expression and Thought

10

possibility of a meaningful metaphor relies on a context within which the language is embedded, such as described by Wittgenstein in a later work [3] in terms of *language-games*. A language-game on Wittgenstein's account brings "into prominence the fact that the speaking of language is part of an activity, or of a life-form" [3, §23]. As Monk [8, p. 330] explains, the purpose of language-games is "to free ourselves from the philosophical confusions that result from considering language in isolation from its place in the 'stream of life'".

These descriptions of Wittgenstein's approach reflect a broader polemic against a position that assumes we can somehow view things *sub specie aeterni*. Wittgenstein viewed such perspectives as negligent of one's own, necessarily earth-bound, position. In a note written to Sraffa in 1935, he describes the irritation caused by the thinking of "Cambridge people", that he formulates as follows: "Here are people who try to speak in a queer way 'impartially' about things, they pretend to be able to slip out of their own skins and they speak as though they could understand everybody's feelings, wishes, tendencies etc." [9, p. 235 n. 7]. In line with this perspective, the notion of a language-game evokes a sense in which, understanding language requires some sort of involvement in it. It is the connection with a game that draws this out, for we can only understand a game (how it is played, what its rules are, what significance it has) through engaging with it in someway. We can no more view our language from an objective perspective than we can *slip out of our skin*. This argument provides some basis to the centrality of context for metaphor, because the last metaphor includes translatable qualities (in terms of seeing things from the point of view of another), but it also has other qualities that make sense from the perspective of an embodied person. Simply: if you've never had skin, can you really understand the *ick* factor that comes when you think in more detail about what it would be like to slip out of it. Let alone to slip *into* the skin of another.

Let us consider another example (which formed part of the title for the first incarnation of this paper): *to find your feet*. In a very general sense the metaphor points to the sense of finding ones way around, or getting to know how things work, where things are, or to familiarise yourself with something in either general or specific terms. The literal meaning makes little sense, since someone with feet and legs will find their feet at the end of their legs where they always have been.[2] In this metaphor, I suggest that this *your* is embedded, meaning*ful*; and ineliminable. This does not mean that context is limited to a singular subjective experience. As Barthes and Wittgenstein both describe, our (linguistic) experiences are shared. Even in vastly different experiences there can be found many sorts of overlap. For instance, one person's experience of a rare or unusual illness does not preclude another person (who has not experienced that same illness) from understanding *something* about what it is to be ill. Illness is not unique, though of course each illness may engender a different kind of experience. Nevertheless the experience of illness *per se* is important to understanding the qualitative experience of illness, just as the experience of skin adds a particular quality of understanding the metaphor offered above. This is not to say that *all* understanding is impossible without it, but rather that the understanding will be qualitatively different, as well as more difficult.

From this we arrive at the crux of the argument, which is that the capacity for understanding arises from experience, and more

specifically the very possibility of that experience. As Kant explains [10, B137/138],

> The synthetic unity of consciousness is, therefore, an objective condition of all knowledge. It is not merely a condition that I myself require in knowing an object, but is a condition under which every intuition must stand in order *to become an object for me*. For otherwise, in the absence of this synthesis, the manifold would *not* be united in one consciousness.

To put this another way, the very possibility of experience is wedded to the possibility of my ability *to* experience. For the purposes of this argument, consciousness here can be replaced by understanding, since the possibility for understanding *metaphor* on this account relies on the condition, or capacity *for understanding*. And understanding, wedded as it is to context, and more broadly experience, is poorer if not embodied and embedded. What this means for computational metaphor is our next concern.

3 COMPUTATIONAL METAPHOR

This argument has been offered in defence of a contextual, experiential, semantic understanding of metaphors. We have not, so far, given consideration to the possibility of computational metaphor, and indeed in stating these words my position is already (partly) declared. I do not doubt that there can be such a thing as computational metaphor (just as in [4] I did not doubt the possibility of computational creativity), but once again I offer the caveat that what it would mean to be *successful* in a computational metaphor (hereafter *c-metaphor*) is not going to be simple, and includes indicators such as appropriateness or even what is acceptable, but more than this it includes the issue of judgement.

Returning to Barthes, we have the question of whether a metaphor stands in judgement on its own, or whether we also judge its origin and what we think it represents. For instance, if I write here about the experience I had this morning drinking coffee, and I want to do this because I want you to know that the coffee I drank improved my mood and my experience of writing this paper, then I would do this because I wanted you to know something(s) about me. This includes things about my mood, my preference for coffee in the morning(s), my experience of writing this paper, and of all the combinations that these elements produce. In so doing my primary motive would not be that you should know something about coffee separate to me and to my experience, especially as I as author chose this example purposefully. Instead, I would want you to know something about *me*. This is no different to conversations that happen about coffee outside of an academic paper. Of course, not all use of words either inspires, requires, or expects this sort of meaning (which is why I think that Barthes is right to be suspicious of the individuality of the author-god), but in this case, as in many other cases, the individual here (me) wants the reader (you) to know something about my experience of the world. If I use a metaphor to illustrate this, say, *this morning's shot of coffee,* then I would highlight both the literal size of the coffee (espresso sized, akin to a shot-sized measure of alcohol), as well as the medicinal quality of having my shot of caffeine. In this way I am pointing to my experience of coffee more generally and in a way that I hope would be familiar to you the reader. Nevertheless I would not want to divorce this metaphor, nor the description that came before it, from my own personal experience this morning. Not because I am an egomaniacal author with god delusions, but

---

[2] In exceptional circumstances, for instance because of a neurological disorder, or an impairment of proprioception, we can imagine someone experiencing a sense of not knowing where their feet are (or even that their feet are their own, rather like in *alien hand syndrome*).

because in the use of a personal experience I quite liked the idea you might understand *me* as a result. Which brings us back to Wittgenstein.

The aphorism at 6.54 [1] does in fact end with an analogy about a ladder, and it's worth a little more consideration:

> My propositions serve as elucidations in the following way: anyone who understands me eventually recognizes them as nonsensical, when he has used them—as steps—to climb up beyond them. (He must, so to speak, throw away the ladder after he has climbed up it.). He must transcend these propositions, and then he will see the world aright. What we cannot speak about we must pass over in silence.

What is particularly interesting about this metaphor is that it is preceded by that word *me*. My argument on this is that, rather like my description of the coffee, the metaphor offered by Wittgenstein cannot be divorced from the author. This is not to say that the interpretations that arise from the text must therefore be ordained by the Wittgenstein-god (since this is both unlikely as well as unnecessary), but rather that the experience that Wittgenstein had with the text, and with the ideas and metaphors he offers, should instead be part of the rich interpretative experience that comes from reading those words. This includes the image of the ladder and all that it might represent. Especially if you've read a lot of Wittgenstein.

Which brings us to c-metaphors. While these can of course satisfy some requirements of metaphor, including claims to novelty, utility, new aspects on the familiar, these descriptors are judged according to a context external to the computer's own capacity, and do in fact follow our own values. Added to which, these values (e.g. of novelty) and utility can contrast with other features of success, for instance, understanding what is trying to be communicated. What, for instance, would a program want to communicate and why? What would a program know of coffee, of skin, of ladders?

In simple terms, do we value a novel metaphor if we do not believe a person (or, in this case, a program) has any idea—including understanding or experience—of the individual components, let alone the comparison being drawn? If, for instance, I had offered the metaphor about coffee to you over lunch, and you happened to know me well enough to know I do not in fact drink coffee,[3] then some value of the metaphor may be lost or at least compromised. We expect that metaphors that reflect an experience have at least some basis in the user's experience otherwise they lose their potency as a basis for communication (as opposed to just literary word play).

This follows especially for unusual or bizarre comparisons that make sense where we look for or expect sense, but not where we might expect little sense (for instance in the babbling of a very young infant). Kingsey Amis' description of a hangover in *Lucky Jim* is one such example, and (to my mind) one of the finest:

> Dixon was alive again. Consciousness was upon him before he could get out of the way; not for him the slow, gracious wandering from the halls of sleep, but a summary, forcible ejection. He lay sprawled, too wicked to move, spewed up like a broken spider-crab on the tarry shingle of the morning. The light did him harm, but not as much as looking at things did; he

---

[3] In fact I do drink coffee, but in a thought experiment anything is possible.

> resolved, having done it once, never to move his eyeballs again. A dusty thudding in his head made the scene before him beat like a pulse. His mouth had been used as a latrine by some small creature of the night, and then as its mausoleum. During the night, too, he'd somehow been on a cross-country run and then been expertly beaten up by secret police. He felt bad. [11]

My faith in this description of a hangover is partly borne out by my own experiences, yet had I not had those, then it would be based in a judgement of the author's, or at least the character's own knowledge, and here it requires not only that we understand the words, but that we understand them meaningfully. The above description by Jim is what it is to have a hangover in his view, as perhaps for Amis, and in terms of the rest of the novel, the description is in kilter. We can of course measure the success of a metaphor based on content, or according to any number of rules, whether these are pre or post hoc, amendable, or alterable, but we can also accept as well as reject metaphors based on context and expectation of meaning, which includes both judgement and bias. If the description of the hangover above had come from someone that you knew to be teetotal, you might still accept its accuracy as a measure of success, but again, the value of the metaphor might be compromised.

If this seems arbitrary or even unfair, I would be inclined to agree. But it's no more arbitrary or unfair than the decisions or processes by which terms either become or cease to be colloquial, slang or popular. What is considered either appropriate or acceptable in ordinary language is also tricky, including where highly creative language-use can muddy the waters of ordinary language substantially (not least where profanities are concerned). It is also partly because of the role that *expectation of meaning* creates. As I discuss above and elsewhere [4] [5], meaningful language-games in Wittgenstein's terms require not only a successful meeting of *rules*, but also a willing on the part of participants to *recognise* other speakers as meaningful language-users. In the case of proposition 6.54 above, it is precisely because scholars expect meaning to be found, that the search for a meaning is considered worthwhile.

C-metaphor construction, interpretation, use, and so on, is not impossible or even unlikely. Whether these metaphors are accepted, adopted or even considered worth paying attention to, however, remains to be seen. Even if the c-metaphor is interesting or impressive, this does not strike me as any more meaningful than when a very small child stumbles across a successful metaphor without really understanding the words or the implications of the word order. This is not to say that they absolutely did not understand, but then again, this is easier to resolve with a program than with a small child, since children do become meaningful language-users.

Where language-use is disembodied and decontextualized, the concept of the language-game makes little meaningful sense. Indeed the metaphor of the *game* is particularly helpful, since it points to the ideas of participation and mimicry. Both are key in the learning and using of language in a meaningful way. As a result, we may not accept a metaphor as creative or even useful if we do not believe the person (or program) has any idea (whether meaningful understanding or experience) of the individual components, let alone the comparison being drawn. Just as we might have doubts about the non-coffee drinker's, or teetotaller's use of certain metaphors about either tea or alcohol. This is not to say we'd necessarily reject the metaphor, but only that we may doubt the success of the utterer or even of the uttered as a result.

4 CONCLUSION

This paper has sketched out an argument about metaphor, which remains in its infancy but which contains a number of propositions. The first is that for metaphor to be meaningful both context and embodied experience is required. These add colour (experience, meaning) to words, through which we come to understand and interpret the words themselves as well as those who utter them. Where this is missing, a crucial element of communication is thereby also missing. The question thus becomes: if you've not experienced colour, then can you really *understand* the metaphor I've offered above?

The author has not sought to suggest that words cannot have a meaning without context. Indeed there are many examples of this in all kinds of places (including on walls). Nor is it the argument that all words that are spoken or written must have an individual intention towards a particular meaning. There is sufficient evidence against such a claim, and Barthes' discussion of the author-god provides some sense of this. The author also finds it acceptable to say that language, at least in terms of signs, can be manipulated without a language-user, though I rather agree with Searle on this point that this can be described in terms of syntax rather than semantics [12].

Instead the author has sought to show that the expectation of a context, an intention, or a speaker is central to finding meaning in words, and particularly in a metaphor or other creative language. I bet you imagined the author as someone who drinks coffee at least once during the reading of this paper, and if you did then you have begun to understand *me*, or at least me as coffee-drinker. Of course this assumes you know about coffee, and have imagination, but I'm happy to assume this about the reader, and to imagine what it might be to be *you*.

ACKNOWLEDGEMENTS

REFERENCES

[1] L. Wittgenstein, *Tractatus Logico-Philosophicus* (Trans. D. Pears & B. McGuinness). London: Routledge, 1974.

[2] C. Diamond, Ethics, Imagination and the Method of Wittgenstein's Tractatus. In A. M. Crary & R. J. Read (Eds.), *The New Wittgenstein* (pp. 149-173). London: Routledge, 2000.

[3] L. Wittgenstein, *Philosophical Investigations*. Oxford: Blackwell (Trans. G. E. M. Anscombe), 2001.

[4] Y. J. Erden, Could a created being ever be creative? Some philosophical remarks on creativity and AI development, *Minds and Machines*, 20(3): 349-362, 2010.

[5] Y. J. Erden, The 'simple-minded' metaphor: Why the brain is *not* a computer, via a defence of Searle, AISB Proceedings, 2013.

[6] G. Priest, *Beyond the Limits of Thought*. Oxford: Clarendon Press, 2002.

[7] R. Barthes, The Death of the Author, (Trans. Richard Howard), New York: Hill and Wang, 1977. Available at: http://ww2.valdosta.edu/~tthompson/ppts/3060/spring2013/death_authorbarthes.pdf (Accessed 20/03/15)

[8] R. Monk, *Ludwig Wittgenstein: The Duty of Genius*. London: Vintage, 1990.

[9] B. McGuinness, What Wittgenstein got from Sraffa. In G. Chiodi & L. Ditta (Eds.), *Sraffa or an Alternative Economics*. Basingstoke: Palgrave Macmillan, 2007.

[10] I. Kant, *Critique of Pure Reason* (trans. N. K. Smith), London: Macmillan press, 1933 (2nd ed.).

[11] K. Amis, *Lucky Jim*. London: Penguin, 1954.

[12] J. Searle, *The Rediscovery of the Mind*. Cambridge, Massachusetts: M.I.T. Press, 1992.

Automatic Metaphor-Interpretation in the Framework of Structural Semantics

Christian J. Feldbacher[1]

Abstract. Given that metaphors can be important parts of arguments and that the common methods for evaluating literal claims and arguments are not (directly) applicable to metaphorical ones, several questions arise: In which way are metaphors important? How do metaphorical premises of an argument support its conclusion? What is an adequate evaluation procedure for metaphorical claims and arguments? In this paper we will give answers especially to the first and second question and indicate how an answer to the third question might look like. Metaphors in arguments—so our analysis—introduce some very general assumptions about the domain of investigation and these general assumptions—spelled out explicitly—are in support of the conclusion of the argument. To render our analysis more precisely we will outline an implementation of automatic metaphor recognition and interpretation with the help of structural semantics. By applying such an implementation it is aimed at reducing the question of evaluation to that one of evaluating by logical or probabilistic means literal arguments.

Keywords: metaphorical argumentation, automatic metaphor recognition, automatic metaphor interpretation, structural semantics

1 Objective

"Religious beliefs are viruses of the mind."—this is a popular metaphor used to argue against religious belief. Metaphors often play an important role in such arguments. They are not only used to attack, e.g., opposing claims, but also to explain why a phenomenon as, e.g., religion has a specific property—here: is so wide spread and firmly established in society as well as significantly involved in cultural processes. In order to analyse such arguments properly, one is in need of an evaluation method for metaphorical arguments. In this paper we are going to sketch a first approach by assuming a reductive stance towards the evaluation of metaphorical arguments. As a reductive stance we propose to first translate metaphorical arguments to literal ones and then analyse them by the ordinary means of logic and probability theory. In especially we are going to sketch our intermediate results on:

- Metaphor recognition
- Metaphor interpretation
- Automation of metaphor recognition and interpretation

2 Analyzing Metaphorical Claims and Arguments

Metaphorical claims and arguments are used quite frequently, even in scientific contexts. The common methods for evaluating literal claims and arguments are not (directly) applicable to metaphorical ones. So one needs an evaluation procedure for metaphorical claims and arguments. Such a procedure may be reduced to classical evaluation procedures for arguments with expressions in literal meaning as follows:

1. Analyze the metaphorical expressions. Outcome of this process is a list of expressions possibly used as metaphors.
2. Find out implicit claims (hidden assumptions). Here we get as outcome a reduced list of such expressions and a list of claims using this expressions.
3. Reconstruct the metaphorical claim or argument. The Outcome of this process is a list of claims containing expressions in literal use only.
4. Evaluate the reconstructed claim or argument using common methods. This is just the standard procedure of evaluating arguments with literally used expressions only.

What is needed for evaluation of metaphorical arguments in the first place, is a method of analyzing and interpreting metaphors which is the main objective of this paper. With 'literal' we mean here the possibly manyfold meaning of an expression that is listed in natural language dictionaries. We intend here only a very rudimentary treatment and incorporation of such meanings, as is present, e.g., in word clouds.

2.1 Simple Accounts of Analyzing Metaphors

Traditional accounts of analyzing metaphors are, e.g., the so-called *substitutional view* (cf. [6] and [3]):

- Metaphors of the form 'X is Y' can be reduced to literal statements of the form 'X is Z, where 'Z' is a literal substitute of 'Y'.
- The metaphor is primarily about X.

and, e.g., the so-called *comparison view* (cf. [4]):

- Metaphors of the form 'X is Y' can be reduced to literal statements of the form 'X is like Y (in being Z)'.
- The metaphor is just as well about X as about Y.

Problems of the substitutional view are to be found in an adequate characterisation of synonymy as is needed in order to figure out adequate substitutivity. Problems of the comparison view lie in the question of how to interpret the likeness-relation between the relata. For this reason more sophisticated accounts were introduced.

2.2 More Sophisticated Accounts of Analyzing Metaphors

A little bit more sophisticated is the so-called *interaction view* of [1]. According to this view, metaphorical usage of language makes some

---

[1] Duesseldorf Center for Logic and Philosophy of Science, Germany, email: christian.feldbacher@uni-duesseldorf.de

implications expressing interactions between the relata. A heuristics to figure out the literal meaning of an expression is as follows:

1. A metaphor of the form 'X is Y' is given.
2. Construct a list of associated commonplaces w.r.t. the secondary subject:
   - C('Y') = $\langle$'Y is Y_1',..., 'Y is Y_m'$\rangle$
3. Construct from CP('Y') a list of implications by transferring the commonplaces of the secondary subject 'Y' to the primary subject 'X' by help of an interpretation function I.
   - I('X','Y') = $\langle$'X is Y_1',..., 'X is Y_m'$\rangle$
4. Select a list of relevant implications from I('X','Y') by means of an appropriate strategy:
   - RI('X','Y') = $\langle$'X is Y_{i_1}',..., 'X is Y_{i_k}'$\rangle$
5. Then 'X is Y_{i_1} and ... and Y_{i_k}' is a possible interpretation (paraphrase) of the metaphor 'X is Y'.

A problem of the interaction view is this: It is not clear how to figure out the commonplaces w.r.t. a subject and then figure out a set of relevant implications. Also the heuristics presented here starts from a situation where metaphors are already identified. So we would like to offer a new account for metaphor recognition and interpretation that makes Black's presupposed concepts more explicit.

To sum up: Problems of the traditional accounts are:

- The substitutional and the comparison view are too vague and non-constructive.
- Black's interaction account is more adequate. But: If automated, it requires a large amount of manual intervention. There is no general method of determining commonplaces and selecting relevant implications.

Our account aims at the following task:

- To develop an adaption of the interaction account that can be automated so that it does only little or not at all require manual intervention.

For this purpose we want to use structural semantics.

3 Automatic Metaphor Interpretation

Automatic metaphor interpretation is a field of linguistics and computer science, concerned with software based analysis of metaphors. There are two main tasks of automatic metaphor interpretation ([cf. 8, p.1029]):

1. Automatic metaphor recognition
2. Automatic metaphor interpretation

Both tasks are closely connected: Simplified speaking, a metaphorical expression in a context is an expression used not in its literal meaning in the context. To give an interpretation of a metaphorical expression is to paraphrase it with expressions used in their literal meanings ([cf. 8]).

3.1 Metaphor Recognition

What does it mean that an expression in a context is not used in its literal meaning?

Definition 1 (very general criterion) *An expression is a metaphorical expression in a context iff*

1. the context is assumed to be semantically perfect and
2. if the expression is used in its literal meaning, then the context is obviously semantically imperfect.

E.g.: 'Achilles was a lion in the battle.'. If we take 'Achilles' to be understood in its literal meaning, i.e. talking about a human, and also 'lion' in its literal meaning, i.e. talking about a non-human animal, then the sentence (context) is obviously wrong (semantically imperfect). Hence, at least one of the expressions is a metaphorical one.

There are three very central notions used in the criterion:

- 'context'
- 'semantical perfectness'
- 'obviousness'

The context in our example was a sentence. But there are many more other types of contexts possible:

- bottom-up, e.g.: arguments, argument hierarchies
- top-down, e.g.: term-forming expressions (e.g. definite descriptions, functors), predicate-forming expressions (e.g. lambda-expressions) etc.

Depending on the context there are different types of semantical perfectness/imperfectness:

- arguments: valid/invalid, strong/weak
- sentences: true/false, adequate/inadequate, etc.
- term-forming expressions: referential/non-referential

With the help of our general characterization we can provide a systematic formal categorization of metaphors:

1. Propositional metaphors. With sub-species, e.g.:
   (a) Identity metaphors: $t_1 = t_2$ ('Juliet is the sun.')
   (b) Monadic predicative metaphors: $P^1(t)$ ('Juliet is brilliant.')
   (c) Polyadic predicative metaphors: $P^n(t_1, \ldots, t_n)$ ('Juliet is Romeos manna.')
   (d) General subjunctive metaphors: $\forall x(Px \rightarrow Qx)$ ('Religions are viruses.')

2. Term-forming metaphors. With sub-species, e.g.:
   (a) Metaphorical names: c ('Romeo' for a charming man)
   (b) Metaphorical functors: $f^n(t_1, \ldots, t_n)$ ('the heart of his beliefs')

One notion still has to be clarified: 'obviousness'. 'Obviousness' seems to be necessary in order to distinguish semantical imperfectness through metaphors from semantical imperfectness in general. E.g., to claim 'All birds can fly.' is just false, not speaking metaphorically. There are different degrees of the obviousness of semantical imperfectness:

D1 Semantical imperfectness through mixing up categories (sometimes also expressed as stating something which is neither true nor false). E.g. 'Colorless green ideas sleep furiously.'
D2 Semantical imperfectness through logical or definitional falsity. E.g. 'Sophia Loren is a star and not a star.' or 'Soldiers are machines.'
D3 Semantical imperfectness through contradicting commonplaces. E.g. 'Achilles was a lion in the battle.'

$$\vdots$$

We assume that obviousness of semantical imperfectness up to the degree D3 is characteristic for metaphors. I.e.: An expression that is not recognizable in a context as a metaphorical expression up to the knowledge of commonplaces counts as being literally used in the context. To illustrate this assumption, let's take our example 'All birds can fly.'!

- '...flies' is defined on a set containing also birds, so there is no mixing up of categories.   D1: passed...
- The claim is neither logically nor definitionally false (the dictionary just states: 'Birds can fly in general.' which doesn't contradict the claim.)   D2: passed...
- The claim also doesn't contradict commonplaces since 'to fly' is even a connotation of 'being a bird'.   D3: passed...

If we consider our example 'Achilles was a lion in the battle.', it turns out that at least one expression is used metaphorically:

- '...is a lion' is defined on a set containing animals (including humans), so there is no mixing up of categories.   D1: passed...
- The claim is not logically false, but definitionally (the dictionary states two opposing characteristics for 'lion' and 'man' (as genus of 'Achilles'), namely 'non-human' and 'human')
                      D2: not passed...

Our choice of semantical imperfectness up to the degree D3 is motivated by the intended automation which is based on dictionaries and semantical networks and not on "world knowledge" in general. Whether this choice suffices to identify adequately a huge set of metaphorical claims remains an empirical question settled by investigations of performances of our heuristics.

The criterion provided here does not allow us to figure out which expression is the metaphorical one. Someone could speak, e.g., about the Achilles of Homer's *Iliad*, fighting bravely the Trojans. But someone could, e.g., according to our analysis speak also about a lion fighting against a rival as bravely as Achilles did. But this kind of ambiguity, as is mentioned, e.g., also in [cf. 2, p.483,p.485], can be resolved by a non-compositional analysis of the statement in question. The question of identifying the target and the source can be decided only with respect to a broader context.

In order to decide this question, we expand our framework and use some important parts of the semiotical theory *structural semantics*, which was invented in 1966 by Algirdas Julien Greimas ([cf. 7, part.V, section on Greimas]). This is no unconventional choice since the framework of structural semantics is commonly used in literary theory for interpreting literature and importantly also for interpreting metaphors in literature.

There are two important notions of structural semantics needed for our automatized metaphor recognition (and later on: interpretation):

- Seme: "The seme is the minimal unit of semantics, whose function is to differentiate significations." ([7, p.317])
- Isotopy: "Greimas defines isotopy as the principle that allows the semantic concatenation of utterances" where the "iterativity (recurrence) of contextual semes, which connect the semantic elements of discourse (sememes), assures its textual homogeneity and coherence." ([7, p.317])

Very simplified speaking one can say that:

- Semes are the minimal semantical units that are mapped to expressions.
- If an expression is used in a text, then the semes of the expression are set.

- The more a seme is set within a text, the more dominant it is in the text (iteration increases dominance).
- The most dominant semes within a text are the isotopes of the text.

Example:

Isotopies: $seme_2$ and $seme_3$
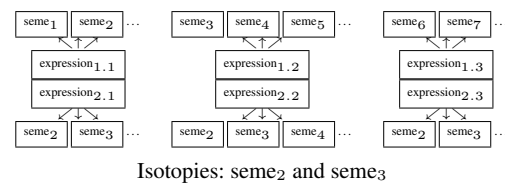
Let's take 'Achilles was a lion in the battle.' with some more context:

Isotopies: 'human', 'fight'
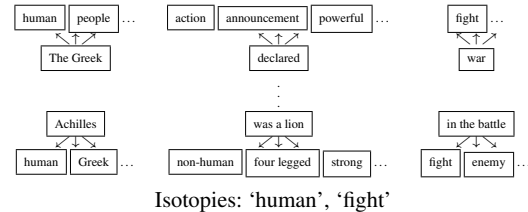
As can be seen, one seme of 'Achilles' is an isotopy, whereas no seme of 'lion' is an isotopy. Since expressions are used normally literally (default), it is likely that metaphorical expressions do not contain isotopies.

We therefore expand the conditions of the criterion for metaphor recognition within the framework of structural semantics:

Definition 2 (more detailed criterion) *An expression is a metaphorical expression in a context iff 1, 2 (of definition 1 above) and:*

3. *No seme of the expression is an isotopy with respect to the overall context. (In comparing expressions one may take the degree of dominance of the expressions' semes for a comparison.)*

The framework of structural semantics is not only useful for the identification of metaphors, but also for their interpretation. In the following we will provide a short sketch of metaphor interpretation in this framework.

3.2 Metaphor Interpretation

Once we have identified metaphors, the question arises of how to paraphrase them in a way such that the paraphrase is non-metaphorical. Just to replace the metaphorical expression by all its semes is inadequate, since this would just make the semantical imperfectness still more obvious (D3⇒D2⇒D1⇒). E.g.:

- If we replace the metaphorical expression 'lion' ...
- ... in the sentence 'Achilles was a lion in the battle.' ...
- ... by its semes 'non-human', 'four legged', 'strong', 'animal' etc. ...
- ... then we end up indeed with a purely literal paraphrase, ...
- ... but on cost of inadequacy:
- 'Achilles was a non-human four legged strong animal in the battle.'

What is needed is some kind of relevance filter, dropping out 'non-human', 'four legged', 'animal' and keeping 'strong'. Here again the *iteration increases dominance* principle of structural semantics is of

some use: The more dominant a seme of a metaphorical expression is within the overall context, the more likely it is to be of relevance.

If the overall context does not increase a seme's degree of dominance, then the seme is less likely to be recognised as a relevant part of a metaphor. And also the other way round: The more dominant a seme is, the easier it is to be recognised as a relevant part of a metaphor. So, for the interpretation of a metaphor one just has to replace the metaphorical expression by the dominant semes to get a literal paraphrase.

3.3   A Fundamental Proviso

Quite common is the point of view that a reductive stance as ours is fundamentally wrong since linguistically and psychologically seen a relation of reduction should be assumed at most the other way round: It is not the literal meaning of an expression we should start of, but a metaphorical one (cf., e.g., [5]). Also Cohen and Margalit claim, e.g., that "it is psychogenetically more illuminating to view literal patterns of word-use as the result of imposing certain restrictions on metaphorical ones, than to view metaphorical patterns as the results of removing certain restrictions from literal ones" ([2, p.470]). Heading into this direction by arguing against the possibility of reducing metaphorical expressions to literal ones, Cohen and Margalit argue as follows—[cf. 2, p.471] (simplified and slightly changed):

1. The meaning of a complex expression is determined by the meaning of its components alone, where the meanings of the basic components are described in dictionaries.
(Principle of compositional semantics)
2. Hence: The meaning of a metaphorical expression is either described in a dictionary directly or is determined by meanings of its components described in a dictionary.   (1)
3. Dictionaries usually record the current use of expressions whereas metaphors are usually innovative, i.e. an expression's metaphorical usage is new.   (general assumption)
4. Hence: The meaning of a metaphorical expression is neither described in a dictionary directly, nor is it determined by—in such a way described—components (otherwise it wouldn't be innovative).   (3)
5. Hence, metaphors cannot be analysed compositionally.   (1, 2–4)

This argument may be seen as counterargument to a reductive stance of metaphors to literal expressions by identifying compositionality with reducibility. Again simplified speaking, Cohen and Margalit propose instead of such a reduction the following analysis—[cf. 2, pp.476ff]: The meaning of an expression is learned inductively by uttering combinations of expressions and taking into account the affirmative or negative responses of trained language users. In doing so one may figure out that, e.g., generally 'shout at me' may go together with 'Peter', but not, e.g., with 'car'. So, we end up with a semantical hypothesis like 'shout' names or describes an action involving as variables a loud tone etc. and is affected, e.g., by the live/non-living variable (according to general usage non-living entities don't shout). Metaphorical usage of 'shout', as, e.g., in 'The car shouted at me.' consists then just in "removing any restrictions in relation to certain variables from the appropriate section or sections of its semantical hypothesis" ([cf. 2, p.482]). So, the psychological relation seems to be as follows:

- Expressions are learned by such combinations and taking into account affirmative or negative feedback.
- Learning of an expression consists in figuring out the relevant variables and putting restrictions on them.

- By this we end up with literal meaning(s) of an expression.
- Speaking in metaphors consists just in relaxing such restrictions again, i.e. in going some steps back in the whole process.

We think that our account is not in contrast to this point of view. Regarding Cohen and Margalit's argument above our approach also denies compositionality, but we still stick to reducibility: According to our theory the correct interpretation of a metaphorical statement is not only based on the meaning of its components alone. Rather it is based on the meaning of its components and the contextually dominant-set semes. By this Cohen and Margalit's claim about the fundamental ambiguity of statements like 'That old man is a baby.' also remains for our approach: "Either its subject is literal and its predicate metaphorical, or vice versa" ([cf. 2, p.483]). Considering the statement alone, 'That old man is a baby.' may be paraphrased adequately by 'That old man behaves like a baby.' or 'That small little thing with this face wrinkled like an old man is a baby.'. But considering it with respect to a context with dominant-set semes as, e.g., the semes of 'experienced', 'wise' etc. in the former and that of 'tiny', 'newborn' etc. in the latter case allows for a disambiguation.

So, to sum up the proviso one may say that our approach also denies the adequacy of compositional reduction, but not that of context-dependent reduction.

3.4   Heuristics for an Automatic Analysis

For automatic metaphor recognition and interpretation in a similar line as described in [10], [9] we used syntactic and semantic databases—at this time only for a text corpus in German (Canoo, Duden, in the future: GermaNet). The flow diagram can be summarized as follows:

- *Basic analysis*
  1. Get the syntactical information of the expressions!   (Canoo)
  2. Transform the expressions into their normal form: Nom.Sg/Inf!   (Canoo)
  3. Extract the semes of the expressions!   (GermaNet)
  4. Extract the connotations of the expressions!   (Duden)

- *Metaphor recognition*
  1. Check whether there are any opposing semes or connotations!
                    (Synonym- and Antonym-Databases)
  2. If so, check which semes are more dominant!
                    (Preceding Analysis)

- *Metaphor interpretation*
  1. Extract the most dominant semes!   (Preceding Analysis)
  2. Transform them into the syntactical form of the metaphorical expression!   (Canoo)
  3. Replace the metaphorical expression by a concatenation of these transformations!

4   Conclusion

In this paper we indicated how two main tasks of theories on metaphors, namely metaphor recognition and metaphor interpretation, may be approached by an automatized analysis. For this purpose the so-called *interaction account* of metaphors served as rough model; we suggested to explicate the key-concepts of this model, i.e.

the concept of 'commonplace' and 'implication', by help of structural semantics: Commonplaces are connections between the semes of an expression and implications are figured out by a dominance operation of the context acting on the metaphorical statement under investigation. Furthermore dominance is operationalized via counting the iteration of semes. The theory is currently implemented into *Perl* for an application on a German text corpus. The implementation is still carried out and it is tried to be expanded on English text corpora too.

References

[1] Max Black. *Models and Metaphors. Studies in Language and Philosophy*. Ithaca: Cornell University Press, 1962.

[2] L. Jonathan Cohen and Avishai Margalit. "The Role of Inductive Reasoning in the Interpretation of Metaphor". In: *Synthese* 21.3–4 (1970), pp.469–487. DOI: 10 . 1007 / BF00484812. URL: http://dx.doi.org/10.1007/ BF00484812.

[3] Raymond W. Gibbs, ed. *The Cambridge Handbook of Metaphor and Thought*. Cambridge: Cambridge University Press, 2008.

[4] Paul Henle. "Metaphor". In: *Language, Thought, and Culture*. Ed. by Paul Henle. Michigan: University of Michigan Press, 1958, pp.173–195.

[5] George Lakoff and Mark Johnson. "Conceptual Metaphor in Everyday Language". English. In: *The Journal of Philosophy* 77.8 (1980), pp. 453–486. ISSN: 0022362X. URL: http: //www.jstor.org/stable/2025464.

[6] Heinrich Lausberg. *Handbuch der literarischen Rhetorik*. München: Steiner, 1973.

[7] Winfried Nöth. *Handbook of Semiotics*. Bloomington: Indiana University Press, 1995.

[8] Ekaterina Shutova. "Automatic Metaphor Interpretation as a Paraphrasing Task". In: *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 2-4, 2010, Los Angeles, California, USA*. Ed. by The Association for Computational Linguistics. Los Angeles: The Association for Computational Linguistics, 2010, pp. 1029–1037. ISBN: 978-1-932432-65-7.

[9] Eric Steinhart. *The Logic of Metaphor: Analogous Parts of Possible Worlds*. Dordrecht: Kluwer Academic Publishers, 2001.

[10] Eric Steinhart and Eva Kittay. "Generating Metaphors from Networs". In: *Approaches to Metaphor*. Ed. by Jaakko Hintikka. Dordrecht: Kluwer Academic Publishers, 1994, pp.41–94.

Metaphorical Minds, Illusory Introspection, and Two Kinds of Analogical Reasoning

Eugen Fischer[1]

Abstract   Introspective conceptions of the mind are inconsistent with recent findings from cognitive and social psychology, but remain intuitive and culturally influential. This paper builds up to a debunking explanation of intuitions which, historically, are at the root of introspective conceptions. The explanation exposes these intuitions as cognitive illusions. It shows that they are devoid of determinate meaning and traces them back to seductive mistakes at the mapping stage of analogical reasoning. The argument employs key principles of the ATT-Meta model of metaphor comprehension and a structure-mapping account of analogical reasoning. The paper argues that, as a default strategy, the comprehension of extended metaphors involves only a very restricted form of analogical inference. It shows how 'full blooded' analogical reasoning with metaphor-transcendent mappings leads to conclusions incapable of metaphorical interpretation through that default strategy. It explains why those transcendent mappings are made, and identifies a previously unrecognised fallacy at the mapping stage of analogical reasoning, the 'metaphor-overextension fallacy'.

1 INTRODUCTION

Intuitive conceptions of the mind, which frequently pass for common sense, credit us with introspective access to, and hence direct knowledge of, a wide range of mental states and processes. These intuitive conceptions have been called into question by several strands of now famous work in social psychology [1, 2], cognitive psychology [3, 4], and cognitive neuroscience [5]. These contributions have forged a new picture of everyday action, decision-making, judgment, and belief-formation: In the absence of determinate prior attitudes or information, people typically perform actions, take decisions and form beliefs due to processes of automatic cognition into which they have little, if any, insight. In many such cases, they then rationalize their actions and beliefs with reasons that do not reflect the factors that moved them. These reasons are hence of little explanatory or predictive value. Instead, rationalisations take up one of several readily available, socially accepted patterns of justification, apparently arbitrarily. The stated reasons might then as well have taken up another pattern, justifying different actions or beliefs. Where this happens, these reasons have only limited justificatory value. It is therefore scarcely an exaggeration to say that, as often as not, when people make up their minds, everything important happens at the level of automatic cognition of which we are largely unaware, and subsequently stated reasons explain nothing and justify little [cp. 6, 7].

To help assess and resolve the manifest tension between this new picture and intuitive introspective conceptions, this paper will prepare the ground for a debunking explanation of relevant

"introspective intuitions", of the kind sometimes sought by one strand of current experimental philosophy [8], known as the 'sources project' [9] or 'cognitive epistemology' [10]. Students of metaphor have prominently suggested these intuitive introspective conceptions are due to unwitting use of visual metaphors [11]. Proceeding from a case-study on four key intuitions from the early modern philosophy of mind, this paper will argue that only a fallacy in analogical reasoning with these conceptual metaphors leads to the intuitions targeted and leads us to give introspection a wider scope than is consistent with the new scientific picture.

The heuristics and biases programme in the psychology of judgment has sought to explain intuitive judgments as outcomes of automatic inferences with heuristic rules which are generally reliable but generate cognitive illusions under specific circumstances [12, 13, 14]. The overtly heuristic character of the rules of analogical reasoning opens up the prospect of an in some ways analogous argument.

To set the stage, we will contrast a default reasoning strategy and a default comprehension strategy: We will consider a default strategy of analogical reasoning, as commonly conceived in cognitive psychology (review: [15]) and build up to a default strategy for motivating and interpreting fresh metaphorical language. We will build up to the latter strategy by integrating notions from the cognitive psychology of analogy and metaphor with psycholinguistic findings about the role of stereotypes in verb comprehension [16, 17], and building on key insights from the ATT-Meta model of metaphor processing [18, 19]. We will argue that a very restricted form of analogical reasoning suffices to build up, e.g., from stereotypical implications of verbs to conceptual metaphors of (roughly) the sort posited in cognitive linguistics (review: [20]) (Section 2).

By reconstructing how the default reasoning strategy can generate four key tenets of an early modern introspective conception of the mind (Section 3), we will then see how, and when, the default reasoning strategy can lead us to cognitive illusions, namely, to illusions of sense: to conclusions which cannot be interpreted with the default comprehension strategy and are therefore liable to lack determinate meaning (Section 4). We will see that this happens the moment more complex analogical inferences employ extensions that 'transcend' the extended mappings properly constitutive of conceptual metaphors. Finally, we will propose an explanation of why these extensions are made (Section 5), i.e., of why perfectly competent speakers come to overextend the conceptual metaphors at issue, namely, in non-intentional analogical inferences [21] which have been found to be involved in problem-solving [22, 23, 24; but cp. 25].

2 TWO STRATEGIES: FULL-BLOODED AND RESTRICTED ANALOGICAL REASONING

Our argument will rely on the distinction between two strategies:

---

[1] School of Politics, Philosophy, Language and Communication Studies, University of East Anglia, Norwich NR4 7TJ, UK. E.Fischer@uea.ac.uk

We now briefly sketch a default strategy for 'full-blooded' analogical reasoning, and then build up to a strategy for motivating and comprehending metaphorical talk, which makes use of a more restricted form of analogical inferencing.

As standardly conceived in cognitive psychology (review: [15]), analogical reasoning about a target domain TD (say, atoms) involves at least three steps: First, a model or source-domain SD (e.g. the solar system) is identified, and knowledge about it is retrieved from memory. Second, model and target are aligned, and elements of the source-model (planets, sun, relations between them: x revolves around y, y attracts x, etc.) are mapped onto elements of the target domain (electrons, nucleus, etc.), subject to semantic and structural constraints: According to influential models of analogical inference (including SME: [26, cp. 27]), we first correlate source- and target-domain elements which are *semantically similar* (which we believe to share properties or stand in the same relations), and then prune these correlations and add new ones by enforcing structural constraints including *1-to-1 mapping* and *parallel connectivity* (when mapping a relation or property onto another, also map their relata or bearers onto each other). Third, the actual inferences are made through *copying with substitution and generation (CWSG)* from a (partial) representation of the source domain SD.

Within the philosophically familiar format of inferences from a set of premises, such *standard analogical (CWSG) inferences* are governed by these three rules: Wherever the premises invoke a SD element which has been mapped onto a TD element,

1. *copy* the representations of relations and relata attached to the SD element, into a set of candidate conclusions about the TD.
2. In the candidates, *substitute* representations of SD relations and relata by representations of TD elements onto which they are mapped.
3. If no such mapping exists, copy the representation of the SD element unchanged into the conclusions ('*generation*').

This default strategy for analogical reasoning contrasts with what I will suggest is a default strategy for motivating and interpreting fresh metaphorical language (*pace* [28]).

According to the ATT-Meta model of metaphor processing, only some of the resources involved in the default strategy for 'full-blooded' analogical (CWSG) reasoning are employed in facilitating metaphorical talk [18, 29]. Two of the model's principles are particularly pertinent for our purposes:

a) Coherent mappings from a source- to a target-domain (conceptual metaphors CM) are built up from single core mappings by a few generic default processes ('vehicle-neutral mapping adjuncts').
b) The mappings obtained with these slender resources are deployed to maximum effect, namely, in interpreting metaphorical uses of expressions which literally stand for 'CM-transcendent' source-domain elements, i.e. for elements which are not mapped by CMs that are built up in this way. Such uses are typically interpreted not by adding further mappings to the conceptual metaphor but by relating the elements 'transcending' it to elements mapped by it.

Many core mappings can be obtained from stereotypical inferences we routinely execute in language comprehension: When interpreting nouns [30] and verbs [16], competent speaker/hearers automatically infer stereotypically associated attributes and consequences, in line with the neo-Gricean I-heuristic: 'Find interpretations that are stereotypical and specific!' [31]. E.g., when people see something happening, they typically know it is happening. Speakers can therefore extend the use of words (e.g., "see") to stand for the stereotypically associated consequence (the subject knows) that hearers will automatically infer, in the absence of explicit indications to the contrary. Such use turns stereotypical into necessary consequences, and defeasible pragmatic into non-defeasible semantic inferences. (You can 'see a kidnapping' without realising what it is, but cannot 'see my point' without knowing what it is.) Such 'pragmatic strengthening' [32] is one of several processes that can endow expressions with metaphorical senses in which they apply in fresh (here: non-visual) contexts [33].

Very elementary automatic analogical inferences [21] can then treat these extensions as cross-domain mappings (here: from the SD of vision to the TD of knowledge) and build up to further, related mappings, which can, in turn, motivate the metaphorical extension of further, related expressions. This happens through generic default operations which unfold, e.g., the conceptual metaphor 'Knowing as Seeing' from the core mapping (here and below, "→" represents mapping, not implication or entailment):

(1)   S sees x → S knows x

These default operations can be conceptualised as the very simplest analogical inferences, namely, analogical inferences which invoke only such a core mapping and generic (e.g. logical) functions and relations which obtain across domains, and hence get mapped onto themselves. These *elementary CWS inferences (ECWS inferences)* involve

(i)   only copying with substitution (CWS),
(ii)   no generation, and
(iii)   employ only core mappings like (1) and 'mappings onto self', which are the first mappings to be made in analogical reasoning (cp. Forbus et al. 1995).

Such elementary inferences can proceed from closed and open sentences. In the latter case, we obtain fresh mappings of relations onto relations. Table 1 gives a particularly simple example, resulting in the fresh mapping

(2)   S does not see x → S does not know x

Table 1. An elementary CWS inference

	SD premise	Operation	TD conclusion
1	¬	Substitution (identical)	¬
2	S sees X	Substitution with (1)	S knows X

Other ECWS inferences yield, e.g.:

(3)   It is possible for S to see x → It is possible for S to know x
(4)   It is not possible for S to see x → It is not possible for S to know x
(5)   X makes it possible for S to see y → X makes it possible for S to know y

(6)   X makes it impossible for S to see y → X makes it impossible for S to know y

According to ATT-Meta, not only logical and modal but also temporal, causal, enabling, and disenabling relations are invoked in generic expansion of core mappings [18, 29]. A core mapping and the further mappings obtainable through ECWS inferences are jointly *'constitutive' of a conceptual metaphor* (here: 'Knowing as Seeing').

Elementary automatic inferences can follow equally automatic stereotypical or semantic inferences. Such brief *inference chains* allow hearers to spontaneously give metaphorical interpretations to further expressions. This motivates the metaphorical extension of these expressions. Consider, e.g., the extension of "beyond my ken" from its literal meaning, 'beyond my range of vision'. When something is beyond someone's ken, he typically cannot get to see it. A *stereotypical inference* hence has it that

(SI)   If X is beyond the ken of S, then S cannot get to see X.

An elementary *analogical inference* (with mapping 4 above) then takes us from the consequent to:

(AI)   S cannot get to know X.

Speakers can extend the use of expressions (here: "X is beyond the ken of S") to stand for the conclusions of such chained inferences ('S cannot get to know X'). A variant of pragmatic strengthening can then make these inferences indefeasible, and the new metaphorical sense conventional. Let's say that the meaning or interpretation derivable through this *two-step default interpretation strategy* is 'induced by the conceptual metaphor CM' that is used for the final analogical inference (*'CM-induced'*).

Where the strategy draws on stereotypical, rather than semantic inferences about the SD, complex expressions will thus acquire as a whole a meaning that is *non-compositional*, i.e., not a function of the meaning, literal or metaphorical, of the expression's constituent parts (here: "beyond", "ken"). Where the strategy employs semantic inferences about the SD, the fresh metaphorical meaning of a complex expression can be regarded as a function (also) of the literal meanings of its constituent parts. In neither case will the former be a function of metaphorical meanings of the latter. These constituents (e.g., "beyond" and "ken") need not have any metaphorical meanings.

In line with the second of our two principles (from ATT-Meta), the metaphorical interpretation of the expression "x is beyond my ken" does not involve reliance on a fresh mapping of the source-domain element 'ken' to the target-domain but rather a chained inference that invokes only a mapping constitutive of the conceptual metaphor. As a default, the kind of analogical reasoning involved in the use and comprehension of metaphors involves only a very restricted range of mappings: the mappings that can be obtained from core mappings through ECWS inference.

3 METAPHORICAL MINDS

As we will now see, introspective conceptions of the mind essentially rely on rather more 'full-blooded' analogical reasoning that (a) involves copying with substitution *and generation* (full CWSG) and (b) invokes both mappings

constitutive of visual metaphors and further mappings that 'transcend' these metaphors. While the terminology varies slightly, seminal early modern texts work with the twin mappings (see, e.g., Fischer [34] on Locke [35]):

Mapping M: visual field → mind
Mapping N: eyes → understanding

These mappings cannot be obtained through ECWS inferences from the core mappings of visual cognition metaphors. Nor are they constitutive of other familiar conceptual metaphors that are linguistically realised in pre-philosophical English. To see this, consider the spatial-inclusion metaphor of remembering and thinking-of which is the home of many uses of "the mind": It unfolds from the core

Mapping R: X is inside a space belonging to S → S remembers / thinks of X

This personal space is typically called 'the mind'. The conceptual metaphor thus motivates saying that we 'keep' or 'have' something 'in mind' when we can think of or remember it, that things 'come to mind' when we actually think of them, and that they 'slip' or (archaically) 'go from our mind' when we forget, temporarily or permanently, etc. [34, pp.41-45]. Where mind-talk is motivated by this metaphor or visual cognition metaphors, "the mind" is used only as part of complex expressions (like "S keeps X in mind", "S's mind was empty" = "S had an empty mind", etc.) whose meanings are not a function of any target-domain meanings of their constituent parts (Section 5). In these contexts "the mind" does not refer to any distinct element of the TD. But mapping M treats the mind as such an element. Hence none of these familiar metaphors include M.

We will now show that analogical reasoning with visual cognition metaphors can take us to the key tenets of classical introspective conceptions of the mind when – and only when – it employs these further mappings which 'transcend' these familiar cognition metaphors [10, 36]. Relevant visual cognition metaphors include the metaphor 'Knowing as Seeing' discussed above (Section 2) and the metaphor 'Thinking-about as Looking-at' which motivates metaphorical talk of 'looking hard at the problem', 'looking at the issue from different angles', or 'looking at the options available'. These conceptual metaphors were extended by adding mappings M and N to them.

Relevant analogical (CWSG) inferences then proceed from source-domain truisms, as in Table 2:

Table 2. A CWSG inference with transcendent mapping

	SD premise	Operation	TD conclusion
1	S looks at X	Substitution: mapping Looking at → Thinking about	S thinks about X
2	(1) Implies (3-4)	Substitution: identical	(1) Implies (3-4)
3	X before Y	Generation	X before Y
4	Y=eyes(S)	Substitution: mapping N	Y=understanding(S)

We thus obtain (non-identical substitutions underlined, generated elements in italics):

P₁   When we <u>look at</u> things, things are before our <u>eyes</u>.
C₁   When we <u>think about</u> things, things *are before* our <u>understanding</u>.

P$_2$ When we <u>look at</u> things, things are in our <u>visual field</u>.
C$_2$ When we <u>think about</u> things, things *are in* our <u>mind</u>.

P$_3$ Things before our <u>eyes</u> are in our <u>visual field</u>.
C$_3$ Things *before* our <u>understanding</u> *are in* our <u>mind</u>.

P$_4$ When we <u>look at</u> things, we perceive things with our <u>eyes</u>, in our <u>visual field</u>.
C$_4$ When we <u>think about</u> things, we perceive things with our <u>understanding</u>, *in* our <u>mind</u>.

These intuitions generate the spatial relations 'X is before Y' and 'X is in Y' in the TD and jointly transform 'the mind' into a personal space of perception, turn 'the understanding' from a 'faculty [!] of reason, intellect, or understanding' (*Oxford English Dictionary*), into an organ of sense that peers into that space, and grant us quasi-perceptual access to the objects of our own thought – but not others'. (Sometimes, 'the understanding' gets replaced by 'the mind' which then doubles as both a space and an organ of 'inner' perception, in violation of the 1-on-1 mapping constraint.)

Crucially, *only* the new mappings N and M take us through familiar visual metaphors to these intuitions and an introspective conception of the mind. To see this, consider what conclusions we obtain through analogical inferences from the present premises when we do not employ the new fare but make do with mappings constitutive of visual metaphors for knowledge or understanding. We then get different conclusions; these conclusions do not generate any spatial relations in the TD; and when interpreted in line with the default comprehension strategy (Section 2) they do not even faintly suggest that thinking involves the use of any organ or space of 'inner' perception.

Relevant inference from P$_1$ yields

C$_1$*   When we think about things, things are before our eyes.

This has a literal interpretation (which is true: when I think – or do anything else, for that matter – something or other will be in front of my eyes, and sometimes I even think about the very things then in front of me). Crucially, it also has a metaphorical interpretation motivated by the visual metaphor: When something is before my eyes, it is typically easy for me to notice (get to see). Stereotypical inference therefore furnishes the premise for an ECWS inference to the conclusion that it is easy for me to get to know or understand. This yields this interpretation of C$_1$*:

'When we think about things, things are easy to understand'

– perhaps unduly optimistic and not idiomatic, but intelligible.
Similarly, analogical inference without M leads from P$_2$ to

C$_2$*   When we think about things, things are within our ken.

When something is within our ken, it is typically possible for us to see. Again, therefore, stereotypical inference furnishes the premise for an ECWS inference (with mapping 3 above) to a straightforward conclusion:

'When we think about things, we can understand things.'

Since none of the elements P$_3$ refers to are mapped by the conceptual metaphors at issue, analogical inferences with these metaphors cannot be directly made from this premise. However, P$_3$ itself employs phrases which have stereotypical implications

in the source domain of vision: When things are before our eyes, it is easy to see them, and when things are in our visual field, it is at any rate possible for us to see them. ECWS inferences lead from the conclusions of the corresponding stereotypical inferences to an undeniable conclusion:

'When things are easy to understand, we can understand things'.

Finally, analogical inference with visual metaphors but without M and N does not take us much beyond P$_4$: Since "perceive", explained by the *OED* as 'to apprehend with the mind or senses', stands for an epistemic relation that can obtain in both the SD of seeing and the TD of cognition, it initially gets mapped onto, and substituted by, itself. We thus obtain:

C$_4$*   When we think about things, we perceive things with our eyes, in our visual field.

But when we perceive something with our eyes, we see it. This semantic implication provides the basis for analogical inferences with core mappings of visual cognition metaphors, e.g., to the conclusion:

'When we think about things, we understand things.'

("…in our visual field" may be disregarded as redundant: how or where else could we possibly see things?) As in the three previous cases, we obtain a conclusion that, interpreted in line with our default comprehension strategy, does not speak of organs or spaces of inner perception.

To sum up: Analogical reasoning with visual cognition metaphors only gets us from SD truisms (like P$_1$ to P$_4$) to the conclusions (C$_1$ to C$_4$) constitutive of the introspective conception of the mind, if we make use of further mapping (like M and N) which 'transcend' those metaphors.

4 ILLUSIONS OF SENSE

We will now outline how and when the use of these further mappings M and N, which 'transcend' visual and other familiar cognition metaphors, can give rise to a particular kind of cognitive illusion: The moment it employs such 'transcending' mappings, the default strategy for analogical reasoning can systematically take us to conclusions which cannot be interpreted either literally or in line with the default strategy for motivating and interpreting fresh metaphorical talk. Barring semantic rescue through fortuitous other conceptual metaphors or metonymies, etc. these conclusions lack determinate meaning. Where they strike us as perfectly intelligible, we are subject to illusions of sense.

Our first set of conclusions, C$_1$ to C$_4$, is a case in point. In contrast with their starred counterparts, they lack metaphorical interpretations motivated by visual metaphors. They all employ at least one of two phrases we obtain when applying N and M to source-domain truisms: "before our understanding" and "in our mind". Neither has a metaphorical interpretation motivated by visual cognition metaphors: In contrast with the source-domain expression "x is before our eyes" from which it is obtained, "x is before our understanding" has no stereotypical or semantic implications in the visual SD. Hence there is nothing for visual cognition metaphors to map, and our default comprehension strategy of making ECWS inferences with mappings constitutive

of the relevant – here: visual – metaphor, from source-domain implications, gets no grip. The same holds true of "in my mind": In contrast to, say, "within my ken", it has no stereotypical or semantic implications in the source domain of vision that could furnish a premise for subsequent ECWS inference with a mapping constitutive of a visual metaphor. The two key phrases lack metaphorical interpretations motivated by visual metaphors.

They also lack literal interpretations: Today as four hundred years ago, "the understanding" ordinarily refers to a faculty. Faculties cannot be literally placed in spatial relations (like the generated relation 'x is before y'). Hence "before our understanding" cannot be interpreted literally. Below (Section 5), we will consider peculiarities of mind-talk and see that, where it is motivated by spatial or visual metaphors, "the mind" always forms part of complex expressions which have no application in the metaphors' SD and possess non-compositional meanings in TD talk. Where a constituent expression (say, "x is in y") takes "the mind" as an argument, it hence cannot be given a literal interpretation. Since C_1 to C_4 all use at least one of the phrases "before the understanding" and "in the mind", these conclusions lack both a literal interpretation and a metaphorical interpretation motivated by visual metaphors.

Other conceptual metaphors, or metonymies, may come to the semantic rescue: E.g., the core mapping R of the spatial memory metaphor (above) lets us interpret the conclusion C_2 as expressing the truism 'When we think about things, we think of things', even if thinkers will have difficulties coming up with this interpretation as long as they are using mapping M. Alternatively, we can exploit semantic entailments ('perceiving' entails 'knowing') and interpret the first part of C_4, 'When we think about things, we perceive things with our understanding' as saying, 'When we think about things, we get to know things by employing our power of reasoning', though thinkers will be unlikely to come up with this interpretation when they are using mapping N. In the absence of such fortunate coincidences (and prior to exploiting them), thinkers are unable to give determinate meaning and content to conclusions like C_1 to C_4. Subsequent *ad hoc* explications were applied inconsistently, frequently disregarded by their own authors, and fail to provide determinate meanings [34, pp.35-41].

The resulting lack of determinate meaning may be obscured by subjective plausibility: C_1 to C_4 have us posit higher-order relations between mapped and generated relations:

(C₁) *When* we *think about* X, it *is before* our understanding.
(C₂) When we think about X, it *is in* our mind.
(C₃) *When* X *is before* the understanding, it *is in* the mind.
(C₄) *When* an object of thought X *is perceived with* the understanding, it *is before* the understanding and *in* the mind.

Such deeply integrated mappings endow analogical conclusions with high subjective plausibility [37, 38]. Furthermore, the posited framework of higher-order relations facilitates inferences from and to constituent and related claims, despite their lack of determinate meaning. E.g.: If something 'is before our understanding' (whatever that might mean exactly), it 'is in our mind' (whatever that might mean here), and 'we perceive it there with our understanding' (ditto). Thinkers may thus be subject to *illusions of sense*: Since they can make various inferences from and to sentences employing these phrases, they may think that these have a determinate meaning, and that they know it, even though they cannot satisfactorily explain the meaning, or apply the phrases consistently to concrete situations.

In our examples, the lack of determinate meaning is due to the use of 'transcendent' mappings M and N. These mappings have us make substitutions within complex expressions (like "before S's eyes" or "within S's ken") that, as a whole, have stereotypical or semantic implications in the SD (e.g. 'It is possible for S to see x') that are mapped onto the TD ('It is possible for S to understand x') by a mapping constitutive of a conceptual metaphor CM. They have us, e.g., replace 'ken' or 'visual field' by 'mind', and 'eyes' by 'understanding'. These substitutions deprive the overall expression E (say, "x is within the ken of S") of the SD implications that facilitate its CM-induced interpretation in line with our default comprehension strategy (Section 2). In this sense, those mappings are *inconsistent with the CM-induced interpretation of E*.

Once metaphorical uses have become familiar or conventional, their interpretation no longer requires analogical inference [39]. The present inconsistency hence does not prevent the philosophers at issue from correctly interpreting familiar metaphorical uses of, say, "beyond my ken" or any other expression E with a conventionalised metaphorical use. The problem arises rather when we use our default strategy for analogical reasoning, in reasoning from SD premises employing a complex expression E: When we then make simultaneous use of a conceptual metaphor CM and mappings inconsistent with CM-induced interpretation of E that has a non-compositional metaphorical meaning, we will obtain a fresh conclusion that cannot be interpreted in line with our default comprehension strategy for metaphorical talk. I.e., our fresh conclusion will lack a default metaphorical interpretation. By forcing substitutions in the complex expression E, those mappings will simultaneously force generation of relations from the remaining frame, in our case the spatial relations 'x is before y' and 'x is in y'. Where such concrete relations are generated in otherwise more abstract talk (like here), literal interpretation of the resulting conclusions is likely to involve category mistakes precluding it ('idea spatially before the understanding', etc.). Failing 'accidental' semantic rescue, such a fresh conclusion will lack determinate meaning.

We have thus built up to a potentially hard-to-spot fallacy committed at the mapping-stage of analogical reasoning. Let's call it the *'metaphor-overextension fallacy'*. It consists in extending a conceptual metaphor CM (such as, e.g., Knowing-as-Seeing) by adding mappings inconsistent with CM-induced interpretations (like mappings M and N). The rules of analogical (CWSG) inference are then liable to take us from true premises to semantically deficient conclusions. Absent semantic rescue through other conceptual metaphors (or fortuitous metonymy, etc.), they will lead to such conclusions whenever CWSG inferences simultaneously employ mappings constitutive of a conceptual metaphor CM and mappings that are inconsistent with the CM-induced interpretation of a complex expression employed in the premises.

5 EXPLAING THE TRANSCENDENT MAPPINGS

But why should competent thinkers commit this fallacy? At the outset (Section 1), we took note of the basic principles of analogical reasoning, as conceived by the influential structure-

mapping theory [40, 37, 26]. We will now identify some factors due to which these principles have us make these mappings even where they lead us from truisms to nonsense.

In some cases, mapping N is straightforward. The structure-mapping account stipulates that in analogical reasoning, with or without metaphor, we routinely add new mappings, where (i) some relations have already been mapped, (ii) the requirement of parallel connectivity demands that we map their relata, and (iii) the target domain contains suitably related elements [41, 42]. This general mapping-rule leads to mapping N, in inferences from premises such as:

P5   When we look at something, we use our eyes.

The first verb is mapped by the basic mapping of the metaphor *Thinking-about as looking-at*. The next verb, "x uses y", stands for a generic relation that obtains in both the visual SD and the intellectual TD. This relation is hence immediately mapped onto itself [27]. This leaves us looking for an element of the intellectual TD that corresponds to our eyes. The latter are introduced here as a relatum of the *use*-relation, temporally linked to the *looking-at* relation that gets mapped onto *thinking-about*. The requirement of parallel connectivity hence has us look for something we use when we think. Since we then use our wits, reason, intellect, or understanding – different labels for the same faculty – we thus obtain

Mapping N: eyes → understanding

Mutatis mutandis, the same applies to inferences employing other visual metaphors, say, from 'When we see something, we use our eyes' to 'When we understand something, we use our intellect'.

Where mappings are *ad hoc*, i.e. involved only in analogical inferences from specific premises, they are easily disregarded in different contexts where they would lead to semantically deficient conclusions. The persistence of N in inferences to such deficient conclusions as the crucial claims C_1 to C_4 therefore requires further explanation.

Parallel connectivity yields N in analogical reasoning from premises like P5, with the core mappings of different related conceptual metaphors: 'Thinking-about as Looking-at', 'Understanding as Seeing', etc. Like many action- and event-nouns [30], all these verbs are associated with quite complex stereotypes known as 'generalised situation schemas' [16, 17]. These are made up of typical features of the action or event that the verb refers to, of the agents performing the action, and of the 'patients' on which it is performed. These features crucially include instruments typically used in performing the action [43]. The strength of stereotypical association is commonly measured through the 'cloze probability' or frequency with which the relevant concept is used to complete sentences such as:

(1) She was sewing the socks with a _____
(2) The man was arrested by _____
(3) When we look at things, we use our _____
(4) When we think about things, we use our ____

The most frequent responses are (1) 'needle' and (2) 'the police' or 'cops' [17]. And while the cloze frequencies for (3) and (4) have not yet been systematically elicited, readers will have little trouble completing them with (3) 'eyes' and (4) 'brains' or 'minds', 'wits', 'reason', 'intelligence' – early moderns would have said our 'intellect' or 'understanding'. Arguably, just as

'sewing' is associated with the subject-property 'uses a needle', 'looking at' is associated with 'uses his eyes', and 'thinking about' with 'uses his brain / mind/ reason / understanding'.

When we encounter or use a verb, all the concepts belonging to the associated generalised situation schema are activated – irrespective of contextual relevance, and the more swiftly and strongly, the stronger the association is [44]. The more strongly a concept is activated, the more likely it will be used in various cognitive processes. If the subject is engaged in analogical reasoning, the concept is hence more likely to be mapped or generated. Where an action or event designated by a source-domain verb gets mapped onto a target-domain concept, all key elements of the situation schema associated with the verb are hence likely to be mapped or generated. Where the schema associated with the TD verb contains an element that stands in the same relation (say, the instrument-relation) to the TD action as the SD associate to the SD action, the SD associate will be mapped onto the TD associate – regardless of whether that relation actually figures in the premise. Thus 'eyes' get mapped onto 'mind' or 'understanding' even in inferences from premises in which the instrument-relation does not figure, like (P1) 'When we look at things, things are before our eyes.' Enforcing the constraint of 1-on-1 mapping in reasoning that also employs mapping M, of 'visual field' onto 'mind', then leads to the preference of 'understanding' over 'mind' we can observe in early modern texts (cp. [34]).

The case of this second mapping M, is more complex. While the patient property 'x is in the visual field of S' presumably is part of the generalised situation schemas associated with vision verbs including "S sees x" and "S looks at x", the mapping onto 'the mind' can never be obtained simply by enforcing parallel connectivity in mapping from SD to TD of a visual cognition metaphor. It cannot, because 'the mind' does not belong to the target domain of such metaphors. In talk motivated by such metaphors, "the mind" is what I propose to call a 'non-member target term'. In first approximation: While it is used only in talk about the target domain, it does not, in any sense, 'stand for' a distinct element of that domain.

To develop this notion, consider how semantic or stereotypical inferences about the SD followed by elementary analogical inferences from their conclusions (Section 2) can motivate common metaphorical expressions. Take, for instance, "S keeps X in mind", as motivated by the spatial memory metaphor unfolding from Mapping R that is the home of English mind-talk. Here, we begin with a semantic inference in the spatial SD:

(SI1)   When S keeps something x in a space (belonging to him), then X continues to be in the space belonging to S.

A mapping of this temporal relation onto the TD relation 'S continues to think of X' can be generated from the core Mapping R through ECWS inferences (what ATT-Meta calls 'vehicle neutral mapping adjuncts'). Analogical inference with this further mapping takes us from the consequent of (SI1) to

(AI)   S continues to think of X.

According to our default strategy, this would motivate a fresh metaphorical use of the SD expression "S keeps X in his space"; instead, we say "S keeps X in mind". Once the chained inference has motivated metaphorical uses of complex expressions including the words "space belonging to S", the latter get

replaced by "mind", as the new lexical item, e.g., "to keep in mind" is formed.

Mutatis mutandis, the same holds true of mind-talk motivated by visual cognition metaphors. Consider how stereotypical followed by analogical inferences could motivate metaphorical uses of complex expressions containing the expression "visual field": Typically,

(SI$_2$) When something is at the forefront of my visual field, I cannot help looking at it.

(SI$_3$) When something is at the back of my visual field, I don't look at it but am aware of it.

Analogical inference with the mapping 'Thinking-about as Looking-at' leads from the stereotypical conclusion (e.g., 'I cannot help looking at it') to a further conclusion (e.g., 'I cannot help thinking about it'). Inference chaining would motivate saying that something is 'at the forefront of my visual field' when I cannot help thinking about it, or 'at the back of my visual field' when I don't think about it, but am aware of it. ('aware of' is a generic epistemic relation that obtains in both source and target domain, hence gets mapped onto itself, and therefore can figure in ECWS inferences of the sort yielding CM-induced interpretations.) But of course we say, instead, that things are 'at the forefront' or 'back of' our 'mind'. Once the chained inference has motivated metaphorical uses of complex expressions including the words "visual field", the latter get replaced by "mind", as the new lexical item, e.g., "at the forefront of the mind" is formed.

"The mind" thus is a *non-member target term* in this more precise sense: On the one hand, it is used only in talk about the target domain, and is not used in talk or reasoning about the source domain. Within the default strategy for motivating and interpreting metaphorical talk, it is not used in reasoning about the SD but replaces source-domain words only once reasoning about the SD has motivated fresh uses of complex expressions containing those words. (In terms of the ATT-Meta approach, the term can figure in reasoning within the pretence cocoon, and its conclusions about the target domain, but not in statements about the source domain.) Hence "the mind" is a 'target term'.

On the other hand, in the cases at issue it merely replaces source-domain terms ("space", "visual field") in more complex expressions. The resulting expressions (e.g. "S keeps X in mind") can be said to refer to elements of the TD, mainly to relations between subjects and objects of thought or knowledge (e.g. 'S continues to think of X'). When the word "mind" is used as synonym of "intellect", etc. it can be said to be individually used to refer to a further TD element, namely, the faculty of reasoning thinkers may employ in thinking. When it is used in metonymies building on this use ("Two great minds [i.e. people with great cognitive abilities] debated the issue"), "the mind" is used to refer to subjects who stand in the relevant relations. But in the present cases, "the mind" merely figures in expressions that, as a whole, have target-domain meanings that are not a function of any target-domain meanings of their constituents. (Indeed, these constituents need not have any such meanings.) In these cases, the constituent expression "mind" cannot be said to refer to any distinct element of the TD: It then forms part of a complex expression that stands for a relationship between a subject and an object of thought or knowledge (e.g. 'S continues to think of X') but not for any further element distinct from such relations and their relata. Hence "the mind" is here used as a

'non-member term': It is here used in talk about the TD but not to stand for any member or element of that domain.

So why does 'the mind' get treated as a TD element, in analogical reasoning which employs mapping M alongside visual metaphors? An as yet speculative answer points out that this may be facilitated by three factors. First, "the mind" replaces words that stand for source-domain elements and whose literal meaning does influence the literal meaning of the complex expressions they enter in. It is therefore tempting to think that the complex expressions into which "the mind" enters must also have a meaning that is a function of the meaning of their constituent parts, and to look for a referent for the constituent "the mind". Since the word is used only in talk about the intellectual target domains, it is natural to look for this referent in them. And, third, the spatial memory metaphor that is its home and anchor has what we may call a *'generic source domain'*: The 'personal space' figuring in core mapping R can be instantiated by an actual physical space belonging to me, e.g., by the space enclosed by my cranium. Hence with R we can motivate saying that I 'cannot keep everything in the head' (when we cannot remember everything) or that we should try to keep certain things 'out of our head' (when we should not think of them). But the conceptual metaphor is not tied to this or any other specific physical instantiation, and the expression "the mind" is used precisely when no such specific instantiation is invoked. This may have us spontaneously rate the term as more abstract and group it with the more abstract concepts from the intellectual TD, rather than the more concrete concepts from spatial or visual source domains invoked.

Once the crucial mistake of treating 'the mind' as a TD element has been made, standard mapping principles have us map 'visual field' onto it: In a first step, SD elements get mapped onto the TD elements deemed most similar to them (Section 2). Through post-inference replacements in antecedents of inferences like (SI$_1$) to (SI$_3$), the 'mind' appears to be credited will all the abstract features (properties and relations) of delineated spaces (in which things can be kept, etc.) and, more specifically, visual fields (which have forefronts and backs, i.e. depth). Through such apparent attributions, 'visual fields' and 'minds' come to be deemed similar enough to get mapped in the first stage of mapping. The presently relevant premises P$_1$ to P$_4$ do not provide any other relata for 'x is in y', so the mapping does not fall foul of structural constraints, in the second stage.

6 CONCLUSIONS

This paper has distinguished two strategies (Section 2): In line with the ATT-Meta model, it has assumed a default strategy for motivating and interpreting (fresh) metaphorical expressions, which makes do with a very restricted form of analogical reasoning, viz., ECWS inferences from core mappings of conceptual metaphors. In line with structure-mapping accounts of analogy, it assumed a default strategy of analogical reasoning that involves a wider range of mappings and full CWSG inference. We then explored how the latter reasoning strategy can lead us from truisms about the visual SD to conclusions about the intellectual TD that cannot be understood through the former interpretation strategy. In the absence of fortunate coincidences, they lack determinate meaning; embedded in inferential links, they strike us as intelligible, even so (Section 4). These illusions of sense are due to mistakes at the mapping

stage of analogical reasoning, namely to an overextension of conceptual metaphors. We explained their extension through problematic mappings by reference to the psychology of schema activation (mapping N) and the peculiar use of "the mind" as a non-member target term (mapping M) (Section 5). The intuitions traced back to these seductive mistakes at the level of mapping are constitutive of early modern conceptions of the mind as a realm of inner perception (Section 3). We have thus obtained a debunking explanation of intuitions at the root of introspective conceptions of the mind. To the extent to which it goes beyond application of key principles of structure mapping theory, on the one hand, and ATT-Meta, on the other, it remains to be computationally developed and experimentally tested.[2]

REFERENCES

[1] R.E. Nisbett and T.D. Wilson. Telling more than we can know: verbal reports on mental processes, *Psychological Review*, 84: 231-59 (1977).

[2] J.A. Bargh, M. Chen, and L. Burrows. Automaticity of social behaviour: Direct effects of trait constructs and stereotype activation on action, *Journal of Personality and Social Psychology*, 71: 230-244 (1996).

[3] A. Tversky and D. Kahneman. Judgment under uncertainty: heuristics and biases. *Science*, 185: 1124-1131 (1974).

[4] P. Slovic, M. Finucane, E: Peters, and D: MacGregor. The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (eds.), *Intuitive Judgement: Heuristics and Biases*. Cambridge: CUP (2002).

[5] M.S. Gazzaniga. *The Social Brain*. New York: Basic Books (1985).

[6] D.M. Wegner. *The Illusion of Conscious Will*. Cambridge, Mass.: MIT Press (2002).

[7] T.D. Wilson. *Strangers to Ourselves. Discovering the adaptive unconscious*, Cambridge, Mass.: Harvard UP (2002).

[8] J. Knobe, J. and S. Nichols. An experimental philosophy manifesto. In their (eds.), *Experimental Philosophy* (pp. 3-14). Oxford: OUP (2008).

[9] J. Pust. Intuition. In E.N. Zalta (ed.), The Stanford Encyclopedia of Philosophy (Winter 2012 Edition). URL = http://plato.stanford.edu/archives/win2012/entries/intuition/ (2012).

[10] E. Fischer. Philosophical intuitions, heuristics, and metaphors. *Synthese*, 191: 569-606 (2014).

[11] G. Lakoff and M. Johnson. *Philosophy in the Flesh*. New York: Basic Books (1999).

[12] A. Tversky and D. Kahneman. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Review*, 90: 293-315 (1983).

[13] D: Kahneman and S. Frederick. A model of heuristic judgment. In K.J. Holyoak and R. Morrison (eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 267-293). Cambridge: CUP (2005).

[14] D. Kahneman. *Thinking fast and slow*. London: Allen Lane (2011).

[15] K.J. Holyoak. Analogy and relational reasoning. In K.J. Holyoak and R.G. Morrison (eds.), *Oxford Handbook of Thinking and Reasoning* (234-59). New York: OUP (2012).

[16] T.R. Ferretti, K. McRae, and A. Hatherell. Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, 44: 516-547 (2001).

[17] K. McRae, T.R. Ferretti, and I. Amyote. Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, 12: 137-176 (1997).

[18] J.A. Barnden. Metaphor and context: A perspective from artificial intelligence. In A. Musolff & J. Zinken (eds.), *Metaphor and Discourse* (pp.79-94). Palgrave Macmillan (2009).

[19] J.A. Barnden. Creative metaphor and metaphorical creativity. In

T.R. Besold, M. Schorlemmer & A. Smaill (eds.): *Computational Creativity Research: Towards Creative Machines* (pp. 217-242). New York: Springer (2015).

[20] R.W. Gibbs. Evaluating conceptual metaphor theory. *Discourse Processes, 48*: 529-562 (2011).

[21] S.B. Day and D. Gentner. Non-intentional analogical inference in text-comprehension. *Memory and Cognition*, 35: 39-49 (2007).

[22] L.A. Keefer, M.J. Landau, D. Sullivan and Z.K. Rothschild. Embodied metaphor and abstract problem solving: testing a metaphoric fit hypothesis in the health domain. *Journal of Experimental Social Psychology*, 53: 12-20 (2014).

[23] P.H. Thibodeau and L. Boroditsky. Metaphors we think with: the role of metaphor in reasoning. *PLoS One 6(2)*, e16782, doi:10.1371/journal.pone.0016782 (2011).

[24] P.H. Thibodeau, and L. Boroditsky. Natural language metaphors covertly influence reasoning. *PLoS One 8(1):* e52961. doi:10.1371/journal.pone.0052961 (2013).

[25] G.J. Steen, W.G. Reijnierse, C. Burgers. When do natural language metaphors influence reasoning? A follow-Up study to Thibodeau and Boroditsky (2013). *PLoS One 9(12):* e113536. doi:10.1371/journal.pone.0113536 (2014).

[26] B. Falkenhainer, K. D. Forbus, and D. Gentner: The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, 41: 1–63 (1989).

[27] K.D. Forbus, D. Gentner, and K. Law. MAC/FAC: a model of similarity-based retrieval. *Cognitive Science, 19*: 141-205 (1995).

[28] P. Wolff and D. Gentner. Structure-mapping in metaphor comprehension. *Cognitive Science*, 35: 1456–1488 (2011).

[29] A.M. Wallington. Systematicity in metaphor and the use of invariant mappings. In: G. Low et al. (eds.), *Researching and Applying Metaphor in the Real World* (pp. 209-244). Amsterdam: John Benjamins (2010).

[30] M. Hare, M. Jones, C. Thomson, S. Kelly, and K. McRae. Activating event knowledge. *Cognition*, 111: 151-67 (2009).

[31] S.C. Levinson. *Presumptive Meanings. The Theory of Generalized Conversational Implicature*. Cambridge, Mass.: MIT Press (2000).

[32] E.C. Traugott. The rise of epistemic meanings in English: example of subjectification in semantic change. *Language*, 65: 31-55 (1989).

[33] E.C. Traugott and R.B. Dasher. *Regularity in Semantic Change*, Cambridge: CUP (2005).

[34] E. Fischer. *Philosophical Delusion and its Therapy*. New York: Routledge (2011).

[35] J. Locke. *An Essay Concerning Human Understanding*, 4th ed. Ed. P. Nidditch. Oxford: Clarendon Press (1700/1975).

[36] E. Fischer. Mind the metaphor! A systematic fallacy in analogical reasoning. *Analysis*, 75: 67-77 (2015).

[37] D. Gentner, M. Ratterman, and K. Forbus. The roles of similarity in transfer: separating retrievability from inferential soundness. *Cognitive Psychology*, 25: 527-75 (1993).

[38] M.E. Lassaline. Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22: 754-70 (1996).

[39] Bowdle, B. and Gentner, D. The career of metaphor. *Psychological Review*, 112: 193-216 (2005).

[40] D. Gentner. Structure mapping: a theoretical framework for analogy. *Cognitive Science*, 7: 155-70 (1983).

[41] D. Gentner and A. Markman. Structure mapping in analogy and similarity. *American Psychologist*, 52: 45-56 (1997).

[42] D. Gentner and A. Markman. Defining structural similarity. *Journal of Cognitive Science*, 6: 1-20 (2005).

[43] M. Harmon-Vukić, S. Guéraud, K.A. Lassonde, and E.J. O'Brien. The activation and instantiation of instrumental inferences. *Discourse Processes*, 46: 467-90 (2009).

[44] R. Giora. *On our Mind. Salience, context, and figurative meaning*. Oxford: OUP (2003).

Metaphors in Theory of Information.

Why They Capture Our Concepts and

Undertakings

Marek Hetmański

Abstract. Metaphors are common in various types of discourse; even natural sciences are engaged with the figurative way of expression mostly characteristic of the humanities. They are also suited, to an astonishing extent, to the exact, strict and formal theories of information, as has been presented in the first part of the paper, on the example of the Shannon & Weaver's Mathematical Theory of Communication. The metaphoric entanglement of the information category shows that its commonsensical and figurative conceptualization is unavoidable. Nevertheless, it also opens certain crucial questions concerning the ways of conceptualizing the probable, uncertain events which happen in the course of communication and deciding.

1 COGNITIVE TOOL

Metaphors are both linguistic and rhetoric means for making analogies between different domains of things. They facilitate the understanding of a complex, obscure, or unfamiliar domain of things, processes, and events through reference to another – one that is more concrete, familiar and comprehensible. Metaphors traditionally function as verbal expressions and utterances of particularly suggestive and pervasive power. They mainly operate as linguistic tools useful in conceiving and describing the world not only in literature but also in science, where they have been manifest and useful throughout the history of science.

But metaphors are not merely verbal in their nature, they are not limited to engaging only the linguistic or communicative competences and faculties of their users. They express deep and complex human mental states and ways of thinking, which are the crucial backdrop for these figurative expressions. Specifically, the nature of metaphors is *conceptual* rather than exclusively verbal - as it is commonly but misleadingly conceived and as is widely investigated and advocated in the theories of *cognitive metaphor* (see [1, 4, 5, 6, 7]). By comparing two different things, processes or events (the subject domains – source and target) with regard to one important aspect, i.e. saying that X *is* (is *like*) Y, metaphor helps to perceive, imagine, and understand one thing (target) in terms of another

(source). Although it is expressed in an expressive, concise way, it is in fact a product of *image schemas* (*conceptual* frameworks) underlying said verbal expression. The frameworks which constitute the agent's mind are sensory-motor in their nature, encompassing such abstract and universal elements as: (1) time and space correlations, (2) before-after things sequences, (3) top-down and/or bottom-up directions, (4) horizontal and/or vertical orientations, as well as the agent's (5) behavioral patterns of movement, manipulation and control. These frameworks organize the agent's experience, be it of his/her immediate environment or the furthest expanses of the universe. Notably, image schemas are especially helpful in trying to envisage the possible, probable or entirely random situations, when planning and predicting the agent's future activities becomes crucial. This has important consequences both in terms of mental and practical aspects of metaphoric discourse. As metaphors shape and guide the agent's behavior in specific directions, they not only explain (as one can obviously expect) that which is is metaphorically expressed, but also unexpectedly hide or obscure is the actual content of the metaphoric thinking. "[A] metaphorical concept can keep us from focusing on other aspects of the concept that are inconsistent with that metaphor" [2]. These somewhat paradoxical consequences will be more closely examined when we consider the metaphoric nature of probable states (Section 3).

2 METAPHORS OF INFORMATION

Claude Shannon and Warren Weaver's Mathematical Theory of Communication [8] is an example of metaphoric thinking engaged in the abstract domain of communication. The linguistic aspect merely implicitly accompanies that which is explicitly (formally, quantitatively) stated. The authors admit using the word *communication* in "a very broad sense to include all of the procedures by which one mind affects another", or as they later specify, "in fact all human behavior (…) one which would include the procedures by means of which one mechanism (…) affects another mechanism" [8]. It is a very broad and general depiction of communication. The examples of communicational mechanisms include not only oral and written speech but also music, theater, pictorial art,

---

[1] Marie Curie-Sklodowska University, Lublin, Poland, email: marek.hetmanski@poczta.umcs.lublin.pl

television, and ballet as well as a guided missile weapons system; all of the above employ procedures of sending, transmitting and processing signals that change the states of the communication processes. However "the language of this memorandum," as Shannon and Weaver relate to their paper, "will often appear to refer to the special, but still broad and important, field of the communication of speech" [8] whereby it aspires to account for all of the above examples of communication. The authors' intention has had certain consequences affecting both their own and other researchers' understanding of information.

The subject of communication as such is considered at three levels: (1) technical – consisting in matching specific signals and symbols while transmitting them during the communication process; (2) semantic – consisting in finding "how precisely do the transmitted symbols convey the desired meaning?"; and (3) pragmatic – "how effectively does the received meaning affect conduct in the desired way?". The last two levels are crucial in that they concern *changes* which communication may bring about, namely, "the success with which the meaning conveyed to the receiver leads to the desired conduct on his part". The essence of communication, including the transfer of signals, lies in changes experienced by the agents involved; information is a function of these changes. The meaning of those signals, analyzed at the semantic and pragmatic levels, is not their main characteristics. It is a relative feature of the transmitted signals and depends on particular sender/receiver intentions. But it is only in ordinary thinking that meaning is identified with information and a particular message having a content. From the point of view of mathematical theory of communication the above statement is misleading. "In particular," say the authors, "*information* must not be confused with meaning" [8]. Two messages – one of which is meaningful and the other completely nonsense – can be formally equivalent and regarded as carrying the same amount of information, no matter the things and situations they refer to.

By adopting the cognitive theory of metaphor as the theoretical background, it is possible to identify in the authors' paper certain crucial elements constituting the structure of each conceptual metaphor. Firstly, there is the *target domain* consisting of the following elements: (1) probable states of events which constitute signals (called "source"); (2) an abstract place/space where signals are transmitted ("channel"); (3) random disturbances of signals as well as interferences between the same and other elements of the channel ("noise"); (4) a way in which signals are organized into a message ("code"); (5) an effective (despite entropy) way of transmitting signals ("redundancy"); and finally, (6) transmission of signals with minimal dispersion to prevent loss of information.

To explain what the above abstract elements (characteristics of any communication) are, Shannon and Weaver provide many *analogies* with empirical and concrete phenomena and situations derived from instances of human communication. They compose a "story" explaining in detail what the subject matter of their paper/report is. In doing so, they constitute a *source domain* consisting of the following, plainly described, consequtive elements: (1) physical signals constituting the message (the news); (2) voice, writing, signals of the nervous system, all of which are constituents of the medium in which transmission takes place; (3) audible sounds or visible seen (e.g. in analog telephony or television) which disturb the

process of communication; (4) language and alphabetic coding ; (5) linguistic and literary styles which help to organize a system of signs into a message; and finally, (6) the actual act of communication. By using self-explanatory and simple analogies to everyday events and situations, the authors try to grasp the essence of information. They do it in a metaphorical – indirect rather than strict or formal – way, which helps them to home in on the general nature of information. But metaphoric understanding of information is neither exclusive nor even dominant over the formal conception of the same. It takes place, so to speak, spontaneously, in accord with ordinary language rules; it shapes the theory in a specific way, leaving on it a remarkable mark. Summarizing their theory of information/communication, Shannon and Weaver write in a tellingly metaphoric way: "An engineering communication theory is just like a very proper an discreet girl accepting your telegram. She pays no attention to the meaning whether it be sad, or joyous, or embarrassing. But she must be prepared to deal with all that come to her desk" [8]. They suggest, in other words, that their conception of information has universal meaning what they express nevertheless through the metaphoric words. Presenting information in this phrase as merely a physical thing (telegram coming to desk) by analogy to the *meaning* of message which is always a concrete thing (which they recommend rather to separate from information as such), they unintentionally but inevitably deprive it of its abstract sense, which depends on *probabilistic* nature of information. In that way mathematical theory of communication due to its metaphoric confinement has been involved in methodological situation. The empirical and vivid elements from the source domain affected, if not dominated, characteristics of the target domain

The mentioned metaphorical aspect of the information theory, generally speaking, stems from the model of a communication act in which the speaker puts ideas (as objects) into words (as containers) and sends them (along a conduit, in a channel) to the listener, who then takes the idea/object out of the words/containers, performing all these activities automatically and without difficulty. This simplified model – to which Michael Reddy refers to as the "conduit metaphor" [6] – is very suggestive and effective in explaining both interpersonal and mass communication. We come across its realizations in ordinary thinking as well as in different conceptions and theories attempting to define communication as such. Mathematical theory of communication is partially tailored to the idea which it reciprocally reinforces. The conduit metaphor generally suggests that communication is reasonable, almost effortless, and does not bring about any interpretational problems. But Reddy argues that this reduced and simplified model fails to represent the actual complexity and richness of human communication; it is presumed that only simple examples of transmissions in mass communication can be reduced to the same. Human communication depends on changing the interlocutors' states of mind but not transmitting the thoughts alongside ideal channel. It occurs and takes place in human minds and acts, rather than in language alone. As it is never perfect, aberrations and disturbances are unavoidable, they are not obstacles but rather circumstances of its development and progress. "They are tendencies inherent in the system, which can only be counteracted by continuous effort and by large amounts of verbal interaction" [6]. The real and rich (informative)

model of such communication must consider dynamic changes rather than static and one-way mechanisms.

3 CHOICE OF PROBABLE STATES

How does the metaphoric confinement of information change our understanding of this category? To what extent does it reveal, or obscure, the essence of the same? Shannon and Weaver seem to be aware of all of these problems and consequences, however, they are not overly focused on the figurative aspect of their discourse. Their main proposition is a purely objective, not subjective (i.e. not agent-oriented), conception of communication and communication. Their basic thesis holds that information is *selection* and *choice* made among the probable states caused and demanded by communication. Transmission of signals involves selecting from a set of alternative states at the source and announcing it at the destination. It concerns not so much what really happens (the fact) as what would happen (possibility) during communicating. "[T]his word information in communication theory relates not so much to what you *do* say, as what you *could* say. That is, information is a measure of one's freedom of choice when one selects a message" [8]. During the process of communication, no messages are simply sent, instead signals are chosen, transmitted and selected. Communicating *per se* is altering both the initial and final states of this process, the result of which yields information. It is therefore in line neither with the common (intuitive) understanding of communication, nor with the model of information as the message. The natural conceptual schemas – linear, before-after sequences of things, as well as time-after sequences of events – underlying the mathematical theory of communication are used by their authors unconsciously. The metaphorical effect is caused without any prior intention.

As they mention that "the unit information indicating that in this situation [i.e. transmitting the signals] one has an amount of freedom of choice, in selecting a message" [6], Shannon and Weaver concentrate on the formal nature of the key concept. Grasping its complex, partially counter-intuitive nature demands a specific cognitive ability. They hold that the abstract "amount of freedom of choice" appeals to any type of communication when the agent's choice – no matter who or what it is, a human being or a machine – results in receiving information. To be more specific and understandable, they turn to figurative modes of expression, which ultimately makes the quantitative problem rather complicated, open to metaphoric discourse. Mathematical (probable) interpretation of information conceives it as an act of choice between possibilities with which the agent is confronted. The agent should distinguish among all probable things, events and processes and then act effectively by selecting one of the same. There is no information without choice, if the agent had no choice at all, information would not appear. Selection and choice among the possible states result increased *uncertainty,* which formally characterizes this situation. "Information is, we must steadily remember, a measure of one's freedom of choice, and hence the greater the information, the greater is the uncertainty that the message actually selected is some particular one. Thus greater freedom of choice, greater uncertainty, greater information go hand in hand" [8]. The authors explain that in order not to fall into "the semantic trap" (when one should remember that the word "information" is used in a special, narrowed meaning), one

ought to conceive information as the concept which "measures freedom of choice and hence uncertainty as to what choice has been made".

4 DECISION MAKING

Coping with the probable states of things and situations is a complex task, both cognitively and practically. It demands proper, prior comprehension of what is probability as such and then a subsequent realization of some general intuitions as well as elementary rules. The ambiguous, somehow counter-intuitive (qualitative) and at the same time exact and strict (quantitative) nature of the *concept* of probability is a challenging issue of science and common experience alike. Its scientific and commonsensical meanings are different in some regards and convergent in others. They are all in principle connected with an act of *making decisions* – a situation in which the agent pursues one direction and steers clear of others on the basis of signals/information he or she receives. For this reason, decision making is a communicational act with an informational aspect; on the other hand, any communication is at the same time intrinsically burdened with choice and decision making.

The decision-making mechanism engaged in communication is commonly compared to tossing up (flipping a coin) or betting on randomized games. This evident metaphorical aspect of conceiving what making a choice/decision when faced with a number of probable states is, brings about certain serious interpretational difficulties. Namely, it demands selecting and choosing the proper picture or model from among all the available alternatives (each with its own metaphorical power) of such a situation. And then the chosen model moulds the comprehension of the nature of probability. In such a situation people perceive and define all types of decision making as concrete games such as dice, roulette wheels or other gambling devices, and also in receiving the news – unexpected and astonishing. Empirical examples derived from everyday life dominate people's imagination and understanding of the choices they are obliged to make. At such times, the probability of scientifically-investigated events (e.g. statistics) is important and decisive.

But the very concept of probability has, in principle, two different meanings – statistical (formal, quantitative) and epistemological (psychological, qualitative) – both of which are constantly misread and used interchangeably thus leading to many problems. "Statistical probability was the sole legitimate form of probability, the sole basis for knowledge. Consequently, »statistical probability« – and the associated world of »randomizing devices« – has become a metaphor for epistemological probability" [5]. The mathematical concept is what gave the idea of probability its content and epistemic aspect. Conversely, *epistemological* probability, secondary and derivative to the statistical one, is the result of preferred theoretical interpretation rather than correlations between actual events. In this sense, the formal (mathematical/statistical) aspects serves as the basis for presenting the target – the agent's imagination of probability as well as his/her experience of uncertainty (mental states). In other words, the abstract serves as a metaphor for the concrete.

Regardless of these ambiguities and reciprocal relations (recognizable at the theoretical level), people commonly conceive, and subsequently cope with, probability as a state of their own *beliefs* rather than events or affairs. It so

happens that statistical probability becomes a definition – a convincing metaphor of people's thoughts and actions – affecting the experience of the world and any knowledge one might claim to have about it. Such a metaphor serves the descriptive function of supplying explanation for unstable, unpredictable, unfamiliar cognitive phenomena such as making choices, predictions or decisions under conditions of uncertainty. Besides, to a certain extent, it also plays a rhetoric function of encouraging people to perform particular socio-cognitive acts with the expectation of securing some profits, especially in the context of randomized events and situations. But in either case metaphoric thinking obscures that which it actually aims to reveal and explain. That is why Raymond W. Gibbs recognizes a specific "»paradox of metaphor« in which metaphor is creative, novel, culturally sensitive, and allows us to transcend the mundane while also being rooted in pervasive patterns of bodily experience common to all people" [1]. It is not particularly rare for this simple figurative manner of thinking to change ways in which more complex phenomena such as the probability of events are conceived.

This seemingly contradictory nature of metaphoric thinking would mean that people engaged in the same are really unable to exceed their physically, mentally and culturally entrenched limits, their conceptual schemas. In transcending what is empirically evident (source domain) and consequently entering cognitively into new, more complex intellectual domains (target), agents are confronted with many empirical constrains – gestures, mental and linguistic schemas, and/or social customs and values. They conceptualize complex and abstract phenomena by means of material, practical devices and instruments, which is especially evident in the context of probability. This specific conceptual-instrumental equipment is of particular use when coping with randomness.

Empirical studies on the mentioned problems of probability and information [2] have lead to interesting conclusions which shed some light on the metaphoric confinement of communication and information. Gerd Gigerenzer holds that all types of decision-making, ranging from simple and intuitive to more complex and rational, are based on limited information. It means they all such choices are far from rational where agents would be equipped with complete and reliable knowledge. Indeed, situations of complete information – where an agent would be able to compute all available courses of action and thus make a fully informed choice – are unattainable. Considering possibilities and selecting probabilities is not algorithmic but mostly heuristic. People tend to make *correct choices* (when buying, investing or communicating) more easily and more often when they are faced with relatively few alternatives, otherwise they would be overwhelmed with the extent of analysis necessary during decision-making. This is a strategy which relies on gut feelings, the so called rule of thumb, in other words intuition. "The quality of intuition lies in the intelligence of the unconscious, the ability to know without thinking which rule to rely on in which situation" [2]. Intuition might give the agent a chance to use more discretional ways of expression, which he/she conceives as similar as well as more (or less) probable. In this way metaphoric thinking combines with intuition and helps us to understand complex situations.

The same correlation has been observed and empirically studied by Daniel Kahneman and Amos Tversky [9] in their theory of making decision under uncertainty. They hold that while making a decision or solving practical and cognitive problems, the agent utilizes relatively constant *cognitive biases* which reflect his/her specific, unavoidable cognitive faults and errors. They include intuitive judgements and beliefs which play a particular role in the assessment of random events and their probability. "[P]eople rely on a limited number of heuristic principles which reduce the complex tasks of assessing the probabilities and predicting values to simpler judgmental operations" [9]. In particular, biases such as: (1) not properly identifying representativeness in a sequence of events, (2) excessive ease in evaluating such sequences, and (3) incorrectly settling statistical problems based on an erroneous evaluation of input data, are decisive for the agent's cognitive faculties. There are also others that result from the agent's cognitive inability to conceive probabilities of events. Namely, the agent assumes erroneous representativeness relative to the transfer of qualities or probability from one class of events to another. It is due to his/her incessant *search for similarities* between facts and events, despite their evident dissimilarity. In conditions of such cognitively biased thinking, the agent becomes especially susceptive to any suggestive expressions that strengthen this tendency, which is when the role of metaphors becomes particularly crucial.

5 PRACTICAL CONSEQUENCES

It is worth mentioning that the problem of developing proper metaphoric concepts and models of information and communication, apart from the strictly methodological aspects of the same, has certain practical consequences. Shannon and Weaver did not consider these consequences to be relevant to only the explanatory aspect of metaphoric phrases they have themselves used on occasion. But if the conduit metaphor, implied in their conception, might confuse people, be it experts, theorists and laymen conceiving what information is and how it is communicated, the issue of the metaphoric confinement of the very concept of information acquires significance. It may influence the way people communicate and decisions while selecting and processing signals and information. Indeed, it may induce or even compel them to make wrong choices while sending and receiving various types of messages such as orders, inquiries, requests, the news, pictures, texts etc. Such instances occur in the context of education, public affairs, political domains or mass culture, wherein communication is fundamental. In these sociocultural domains – in their institutions and organizations such as schools, colleges, universities, libraries, cultural, scientific and research centers – metaphoric phrases, definitions and conceptions of information and knowledge are of particular importance. Only metaphors possessing dynamic and probabilistic, rather than static or linear connotations in their source domain can describe processes of knowledge acquisition and communication whose quantitative aspect is information. By appealing to astonishing phenomena, they can adequately anticipate new and unforeseen informational processes and events; their rhetorical impact would thus change the previous, conservative conceptualization. Only such enriched figurative thinking is able to evoke human *creativity* in cognitive, intellectual, social and cultural areas.

The cognitive, or more precisely descriptive role of informational metaphors is largely realised within the

discipline of information and knowledge organisation, which commonly employs the definition of information formulated by the mathematical theory of communication. As was already discussed in [3], metaphors pertaining to various data bases utilised by libraries, offices or governmental bodies, as well as any open (internet) repositories of knowledge, play a significant role in defining ways in which these can be organised and used. Rather than merely describe and model, they also provide opportunities for creation and administration, as well as, most importantly, effective utilisation of the same by various users. Many of the existing metaphors of knowledge organisation employ metaphorical descriptions, many dating back as far as antiquity or the middle ages, which compare accumulated and available knowledge to buildings (towers, libraries), labyrinths, vast open spaces (on land or sea), trees, maps, networks, or rootstalks. Each of the above emphasises the physical and spatial (geometric, linear and finite) characterisation of knowledge which is typically depicted as a complete and perfect source of information. Consequently, any attempts to acquire knowledge, expand it, discover new content, or establish new connections, will be described using metaphors such as juggling, wandering, exploring, leafing through, deciphering, enquiring, responding, etc. Such metaphors will normally emphasise a rather passive and unproblematic use of information gathered in static and invariable deposits and data bases. If such metaphors are to serve the function of directives or recommendations, rather than merely descriptions or models, they are likely to be addressed to persons involved in the creation and management of such resources, and not so much to regular users of knowledge systems. The latter have in recent decades been approached with ever more plentiful metaphorical expressions pertaining to the internet, which predominantly carry either clearly positive or negative cognitive and emotional connotations and relate to repositories of information and processes of researching for the same. If inclined positively, such metaphors employ phrases whose source domain includes such positively charged expressions as surfing, exploring, richness, surprise, enrichment, etc. Otherwise, information and the internet may likely be metaphorically described as junk, smog, excess, boundlessness, impoverishment, threat, etc.

All informational metaphors (regardless of their axiological associations) become significant only if used in such a way that, aside from their obvious function of describing (modelling) the existing knowledge and information resources, they also encourage their addressees to engage in a particular course of cognitive action. Shannon's conclusion that the gist of information refers not to what is, but to what *can be communicated*, constitutes an important methodological directive in constructing metaphorical references to knowledge and information. Namely, they should refer to the cognitive expectations of particular agents and the realistic possibility of their fulfilment, rather than merely ready-made realisations and factuality. To accomplish this, however, it is necessary to have a criterion allowing for a distinction between: (1) real (realised, own) cognitive *needs* of internet users and (2) apparent (imposed, unrealised) cognitive *demands* encountered when using software tools and applications. Such metaphorical expressions – suggestive but free of obtrusive marketing and advertising tricks – should take the form of directives and guidelines, commands and, most importantly, warnings addressed to internet users.

Any metaphors but particularly those functioning as suggestive linguistic expressions have (as dictated by their rhetorical and eristic origin) a considerably persuasive force which is manifested through inspiring specific behaviours. If an informational and communicational metaphor comprises in its source domain expressions and phrases relating to the expected, possible, and likely, rather than exclusively actual and unambiguous cognitive situations, it will be successful in performing its persuasive function. It can then become an instrument shaping the attitudes of the cognitively wealthy rather than just the informatively impoverished. Moreover, a properly structured metaphor of knowledge organisation will facilitate internet users in making decisions and tackling cognitive problems, wherein access to suitable information is the necessary condition of success. By indicating possibilities and likelihoods – hidden behind apparent information, unavailable to software users overly preoccupied with the operation of these instruments – such a metaphor may reveal the full informative value of a cognitive situation and allow its due recognition.

6 CONCLUSION

It has been shown that metaphoric phrases used by Claude Shannon and Warren Weaver in their Mathematical Theory of Communication are only complementary, and not main in describing what information is. Owing to the theory of conceptual metaphor, one can recognize the implicit mental structures underlying such way of conceptualizing. It has also been suggested that informational metaphors might constitute useful instruments in coping with probable states while making decisions.

REFERENCES

[1] R.W. Gibbs, (Ed.), *The Cambridge Handbook of Metaphor and Thought*, Cambridge University Press, Cambridge, 2008.

[2] G. Gigerenzer, *Gut Feelings: The Intelligence of Unconscious*, Penguin Group, London, 2007.

[3] M. Hetmański, "The actual role of metaphors in knowledge organization", in W. Babik (Ed.), *Knowledge Organization in the 21st Century*, Ergon Verlag, vol. 14, 73-79, 2014.

[4] J. Lakoff, and M. Johnson, *Metaphors We Live By*, The University of Chicago Press, Chicago-London, 1980.

[5] A. Ortony, A. (Ed.), *Metaphor and Thought* Cambridge University Press, Cambridge, 2nd ed., 1993.

[6] M. Reddy, "The conduit metaphor: A case of frame conflict in our language about language", in A. Ortony, (Ed.), *Metaphor and Thought*, Cambridge University Press, Cambridge, 164-201, 1993.

[7] D. Ritchie, "Statistical Probability as a Metaphor for Epistemological Probability", *Metaphor and Symbol*, Vol.18, No. 1, 1-12, 2003.

[8] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication* The University of Illinois Press, Urbana, 1948/1964.

[9] A. Tversky and D. Kahneman, "Judgment under uncertainty", in A. Tversky and D. Kahneman, (Ed.), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge, 3-28, 1974/1998.

From Metaphor to Hypertext: an Interplay of Organic and Mechanical Metaphorics in the Context of New Media Discovering

Zuzana Kobíková[1], Jakub Mácha[1]

Abstract. Hypertextual linking of information is one of the basic principles of digital media. We suppose this principle to be discovered in metaphorical thinking with the help of the so-called absolute metaphors. We derive the notion of an absolute metaphor from Hans Blumenberg's metaphorology, and we interpret metaphors according to Max Black's interaction theory. Our aim is to interpret these absolute metaphors as being open to new implications, just as they are open to a pragmatically determined dialectical interaction of organic and mechanical metaphorics. We follow the direction of interactions within these metaphorics in a philosophical attempt to explain the nature of mechanical and organic systems. In particular we will analyse the metaphors 'association is trail' (Bush), 'computer is a clerk' (Engelbart) and 'hypertext is a Xanadu' (Nelson). All these metaphors are both organic and mechanical. That is why we can say that hypertext is both an organic and mechanical system.

'It is reality that awakens possibilities, and nothing would be more perverse than to deny it. Even so, it will always be the same possibilities, in sum or on the average, that go on repeating themselves until a man comes along who does not value the, actuality above the idea. It is he who first gives the new possibilities their meaning: their direction, and he awakens them. But such a man is far from being a simple proposition. Since his ideas, to the extent that they are not idle fantasies, are nothing but realities as yet unborn, he, too, naturally has a sense of reality; but it is a sense of possible reality, and arrives at its goal much more slowly than most people's sense of their real possibilities.'

Robert Musil, *The Man Without Qualities*, ([1], p. 12)

1 INTRODUCTION

It has been convincingly argued (e.g., [2,3,4]) that a metaphor appears often at an outset of scientific discoveries. We can illustrate this statement in the case of the discovery of hypertext. As a nonlinear text with links containing references to other pieces of information, hypertext presents a new form of media, formed through the remediation of a prior, analogue medium of a text. We present how is the outset of this discovery articulated in a figurative way of metaphor and model.

Etymologically speaking, metaphor means a transfer. According to Arendt [5], we need to use a metaphor, when we need to transcend the borders of the real (given) world and then lead

into speculation, (in our case speculation about the as yet non-existing hypertext, which we have no words for yet). A metaphor means, in this sense, a transfer from something imagined into something existing, thus into a material, functional medium. This is possible with the help of so called predicative metaphors based on analogies. We thus interpret the process of the invention of hypertext as a metaphor in the sense of transfer, which bridges the gap between a possible and an existing reality, as suggested in our epigraph from Musil.

We want to show why that figurative thinking is constructive and worthwhile in the discovery of hypertext and its explanation, and for which roles metaphors and models play in the scientific conceptualising of hypertext.

We will suggest that all inventors of hypertext concepts, mentioned in this paper, make up their concepts of hypertext by employing so-called absolute metaphors. This term, from Blumenberg ([6], pp. 62–69), means a background metaphorical complex, or some leading idea, that systematically informs the thinking of individuals and entire epochs by reference to an implicit model, such as a mechanism or an organism.

We will present the concepts of hypertext as systems based on metaphors, which connect organic and mechanical metaphors together. We will show this in detail with the help of the analysis of the following predicative metaphors, which we suppose to be absolute:

- 'association is trail' (Vannevar Bush, 1945) [7],
- 'computer is a clerk' (Douglas Carl Engelbart, 1962) [8],
- 'hypertext is a Xanadu' (Theodor Holm Nelson 1974) [9].

We want to show that above mentioned metaphors of hypertext are not mutually independent. They have evolved from the first one to the third one, as we will show below.

Vannevar Bush (1890–1974) is our first prototype of Musil's man with a sense for a possible reality. Bush started the transfer between a real (unsatisfactory) and a possible (better) feature of a new form of text. We will interpret his memex as a theoretical model developed from the 'association is a trail' metaphor. Bush wanted to improve the way how scientists deal with information. His memex (imagined as a mechanical machine) would archive all the data that a scientist has collected. The memex would link all this information by means of metadata indexing. Bush 'mobilized' his idea by the means of scientific communication. He described himself as a man of the mechanical age [10] and who wanted to address the scientists of the digital age and to encourage them to transfer his theoretical model onto a functional medium.

The Engelbart and Nelson concepts of hypertext are built upon Bush's metaphor.

[1] Dept. of Philosophy, Masaryk University, Arna Nováka 1, 602 00 Brno, Czech Republic. Email: {kobikova, macha}@mail.muni.cz.

We can start with a presentation of this organic and mechanical interplay of metaphorics from a methodological and historical viewpoint.

2 METHODOLOGY

We draw on Hans Blumenberg's metaphorology [6], combined with Max Black's interaction theory of metaphor [11], and his view of how metaphors and models are used in scientific discoveries [2]. Why have we chosen these authors? Black's semantic theory has now been superseded by pragmatic accounts, which rightly accentuate the pragmatic dimension of the metaphor. But we do not need to focus on the actual speech situation. In this paper we focus on metaphors and their implications mentioned in the scientific texts about hypertext.

Blumenberg's metaphorology resembles Lakoff's and Johnson's theory of the conceptual metaphor [12] which has received much more attention in the past few years. But Blumenberg's account is arguably more complex in its historical point of view, which is also our main focus.

Metaphorology is not just another theory of metaphor in our modern sense, i.e. an analysis of the concept of metaphor, but it is an investigation into some prominent instances of this concept. The first aim of metaphorology is to substantiate the existence of the so-called *absolute metaphor* which, hypothetically for the time being, can be considered as a foundational element of philosophical language. According to Blumenberg, absolute metaphors cannot be translated into unambiguous literal language,[2] they are, so to speak, 'resistant' ([6], pp. 3–5). Blumenberg does not however explain why this or that metaphor is absolute. In his historical perspective, a metaphor is absolute if it has resisted being fully translated thus far. This does not exclude the fact that such a metaphor could be fully translated in the future. We suppose that, in a nutshell, a metaphor is absolute (for a given period), if every attempt at its explanation results in another metaphor or analogy.

The fact that an absolute metaphor cannot be translated into literal language – and this is the second step in Blumenberg's project – does not prevent it from replacing or correcting another absolute metaphor. Such transformations take place in history and they are important subjects of metaphorology ([6], p. 3).

For instance, there are a lot of metaphors about the world: 'the world (order) is (like) a machine' (*machina mundi*) or 'the world is clockwork'.

These two metaphors are not mutually independent, as the latter is a certain specification of the former. In numerous quotations from philosophers and scientists, Blumenberg tried to show how the *machina mundi* metaphor has been transformed into the clockwork-metaphor with the dawning of the Enlightenment ([6], pp. 62–69).

In this paper we will focus on two particular metaphors or rather metaphorical themes (which we call 'metaphorics') – on mechanical and organic metaphors, their dialectical interplay and blending when explaining the nature of associative memory, text and hypertext. In order to do so, we follow Blumenberg's need to examine the *consequences* of this or that particular metaphor by various thinkers. A set of non-contradictory consequences of

---

[2] By 'literal language' we mean the unambiguous language of modern science.

a metaphor is what we call, following Black's interactions theory [11], its *interpretation*.

Max Black provides a complex method of interpreting vital, predicative metaphors of the form 'A is B'. The basic idea is that if such an utterance is intended or/and recognized as a metaphor then the literal meaning of 'A' *interacts* with the literal meaning of 'B' resulting into a *metaphorical* meaning 'B' which is hereby being predicated of 'A' The core of this method consists of explaining how these two meanings interact. They do indirectly through so-called implication-complexes or associated implications. An implication-complex is a set of implications predicable to a term. An implication complex A is a set of implications in the form of 'A implies A_i' and an implication-complex B is a set of implications in the form 'B implies B_i'. These implications do not need to be true; they only have to be considered to be true in a given context. The very interaction consists of pairing members of these complexes $f([A_i, B_i])$. The meaning B_i is transformed by f so that it is predicable of A instead of A_i. The function f may stand for an '(a) identity, (b) extension, typically ad hoc, (c) similarity, (d) analogy, or (e) what might be called a metaphorical coupling', (where, as often happenes, the original metaphor implicates subordinated metaphors). ([11], p. 31) Black does not further explicate these terms. For our purposes, we will take *identity, extension, similarity* to be nonfigurative transfers based on a surface similarity. *Analogy* based on a structural similarity and *metaphorical coupling*, based on a subordinate metaphor are, on the other hand, figurative connections of two implications. They are nested metaphors.

Let us illustrate this method with an example of Thomas Hobbes' mechanical metaphor 'Consequence is a train of thoughts'.[3] The implication-complexes, which depend on the context of utterance or reception, might be:

Thomas Hobbes: Consequence is train of thoughts			
Primary subject: consequence	Secondary subject: train of thoughts		
Implications	**Implications**	**Pairing**	**Way of pairing**
consequence is a succession	train implies movement	[succession, movement]	extension
consequence is a link connecting thoughts	train is a link connecting parts	[link, link]	identity
consequence is a causal connection	train connection is mechanic	[causal, mechanic]	extension
consequence is difficult to avoid	train is difficult to stop	[difficult to avoid, difficult to stop]	analogy
Additional implications			
consequence follow logical laws	trains follow timetables	[follows logical laws, follows timetable]	metaphorical coupling

---

[3] 'BY "consequence," or "train," of thoughts I understand that succession of one thought to another which is called, to distinguish it from discourse in words, "mental discourse."
When a man thinketh on anything whatever, his next thought after is not altogether so casual as it seems to be. Not every thought to every thought succeeds indifferently.' ([13], Ch. III, p. 11.) Hobbes' emphasis on a causal connection between thoughts gives us the reason for taking this metaphor to be mechanical.

Table 1. Interpretation of Thomas Hobbes' mechanical metaphor 'Consequence is a train of thoughts'

The first pair is a case of an extension. The concept of a train's movement is extended so that it covers a succession of thoughts. The second pair is a plain identity. The third pair may be a case of an extension as well. The mechanical way of a train's moving is extended to a broadly causal way of our logical thinking functions (or at least, that is what Hobbes believed). The fourth pair seems to involve an analogy, where the difficulty of bringing a train to standstill is analogous with the difficulty of avoiding a derivation of a consequence. The last pair is a case of an analogy, or a metaphorical coupling. Logical laws are analogous to timetables.[4] However, in which respects? They both express regularities – in a train's movement and in our thinking. Or they both have a normative force, i.e. they both prescribe how things ought to be. There are many aspects in which logical laws are like timetables. Here it is a case of a nested metaphor whose interpretation is open-ended. If this is so, then the interpretation of the original metaphor 'Consequence is a train of thoughts' is open-ended as well.

This example shows that (interpretations of) some metaphors are open-ended or unbounded. This means that such metaphors cannot be easily captured by literal paraphrases. They are absolute metaphors in Blumenberg's sense. Black's interaction theory is, thus, rich enough to be used for analysing absolute metaphors. Black's terminology enables us to recursively qualify metaphors as absolute. A metaphor is absolute if its implication-complexes are connected by analogy or a nested metaphor that is absolute too, because organic and mechanical metaphorics interact here.[5]

Black sees every implication-complex supported by a metaphor's secondary subject as a model of the ascription imputed to the primary subject ([11], p. 31) He develops this theory into the so-called theoretical model. (We describe the memex in terms of a theoretical model in Section 4.) Theoretical models resemble the use of metaphors in requiring analogical transfer of a vocabulary. Metaphor and model creating reveal new relationships. But a metaphor operates largely with *commonplace* implications, says Black, but the author of a scientific model must have prior control of a well-knit scientific theory. Systematic complexity of the source of the model and a capacity for analogical development are essential qualities of models. Black cites another philosopher of science, Stephen Toulmin:

'It is in fact a great virtue of a good model that it does suggest further questions, taking us beyond the phenomena from which we began, and tempts us to formulate hypotheses which turn out to be experimentally fertile… Certainly it is this suggestiveness, and systematic deployability, that makes a good model something more than a simple metaphor.' ([14], pp. 38–39)

---

[4] To be sure, Hobbes couldn't have had in mind trains as we have today. But wooden railways were common in England in the 17th century. They were used for transporting coal from mines. The fifth implication most probably wasn't intended by Hobbes. However, this need not stop us interpreting the metaphor beyond its author's intention.

[5] There can be other reasons of unparaphrasability as the impossibility to spell out all the implications in practice (because they are too subtle, or there are infinitely many implications, or the metaphorical theme is too abstract). These reasons are not our concern.

A successful model must be isomorphic with its domain of application. In stretching the language, by which the model is described in such a way as to fit the new domain, we pin our hopes upon the existence of a common structure in both fields. If the hope is fulfilled, there will have been established objective ground for the analogical transfer. We can determine the validity of a given model by checking the extent of its isomorphism with its intended application. In appraising models as good or bad, we can, in principle at least, determine the 'goodness' of their 'fit'.

In the next section we move to some deeper characterizations of mechanical and organic metaphorics from a historical perspective. We introduce the dialectical relationship between these two metaphorics on examples from Plato's, Kant's and Alberti's absolute metaphors.

3 MECHANICAL AND ORGANIC METAPHORICS FROM A HISTORICAL POINT OF VIEW

The mechanical, as well as the organic metaphorics has a long history. Mechanical metaphors are usually expressed in terms like 'mechanism', 'mechanics', 'machine', but also by 'construction'. Organic metaphors are connected with 'organism', 'life', 'vitality', 'generative' and its cognates. Mechanical metaphorics mean often-detached elements, atoms, driven by abstract forces that exhibit certain regularities or laws. Mechanisms are constructed or discovered by a *bottom-up* approach where pieces, elements, atoms are *composed* together to give rise to a complex system. Elements are prior to the whole. Organic metaphors, on the other hand, highlight the priority of the whole over its parts or the priority of a principle over its instantiations. Parts are here only because of the whole, which is more than a composition of its parts. Organic systems are recognized by a *top-down* approach where the whole is *decomposed* into its functional sub-systems.

The main idea, which drives our investigation, is that of a dialectical relationship between organic and mechanical metaphorics. They are interconnected or even entangled into each other. A mechanical explanation is usually insufficient at a certain point or to a certain extent – an absolute metaphor cannot be fully explained. This gap can be filled by an organic explanation. And this is true also the other way around.

Kant sought in his first *Critique* that nature can be explained by mechanical laws which are derived from the forms of our understanding. This explanation turned out to be insufficient in explaining actions of humans as free beings, but even in explaining some objects occurring in nature like living organisms. They have to be explained teleologically by their inner purposiveness. We can better understand a living organism by asking what its purpose is in nature, not by tracing back its mechanism, which defies any mechanical explanation. Teleological (organic) explanations, however, have for Kant only a heuristic, so to say provisional, role by showing us the directions where to look for mechanical explanations.

The opposite direction is also conceivable. Machines are imitations of organic bodies. This is the traditional Aristotelian view of technology as mimesis. Machines are, in some respect, enhanced bodies (e.g. they are stronger or less prone to malfunctioning), they are, in some other respect, deficient (e.g. they lack

intelligence or they are single-purpose). Here is an illustrative passage from Leon Battista Alberti ([15], p. 175):

'Here we need only consider the machine as a form of extremely strong animal with hands, an animal that can move weights in almost the same way as we do ourselves. These machines must therefore have the same extensions of member and muscle that we use when pressing, pushing, pulling, and carrying.'[6]

Machines are conceived here as extensions of human powers, which is something that will be important in the theories of hypertext. Only (human) organisms as opposed to machines can initiate causal claims.

It is typical that mechanical metaphors aim to explain organic systems and *vice versa*. To use Black's terms, mechanical metaphors are nested in the implication-complexes of organic metaphors. We can, thus, use a mechanic explanation within an overall organic system (and *vice versa*). The decision whether one takes or prefers an organic or mechanical vocabulary depends on the communicative intentions of particular authors. Blumenberg calls this a 'pragmatics function of absolute metaphors.'

In the following three sections we will focus on mechanical and organic metaphorics, their dialectical interplay and blending when explaining the nature of memory, text and hypertext. In order to do so, we, following Blumenberg, need to examine the *consequences* of this or that particular metaphor by hypertext thinkers.

4 MEMEX: MECHANISATION OF ORGANIC MEMORY

We begin this section with an analysis and interpretation of the metaphor 'association is a trail', abstracted from Bush's text. We have chosen it because it helps us to understand as the basic metaphor of hypertext. Engelbart and Nelson (subsequent hypertext investigators) further developed their hypertextual systems from the 'association is a trail' metaphor by developing its open implications. From a theoretical point of view, the 'association is a trail' metaphor fulfils our criteria of an absolute metaphor born from an organic and mechanical background metaphorics. In accordance with Bush, we consider an association as organic, connoted with complexity, unpredictability and intricacy. A trail seems to be more mechanical, systematic, better marked, and easier to follow – at least in Bush's overall aim to mechanize human memory.

Let us follow the directions in a dialectical interaction of organic and mechanic metaphorics in the 'association is a trail' metaphor. Bush describes the methods of mechanical, artificial indexing, which he finds inappropriate at first.

'[…] significant attainments become lost in the mass of the inconsequential […] Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing. When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used; one has to have rules as to

which path will locate it, and the rules are cumbersome. Having found one item, moreover, one has to emerge from the system and re-enter on a new path.' ([7] p. 1)

The mechanical way of linking content is insufficient. Therefore Bush finds a solution in the organic quality of an association:

'The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain. It has other characteristics, of course; trails that are not frequently followed are prone to fade. Items are not fully permanent and memory is transitory. Yet the speed of action, the intricacy of trails, the detail of mental pictures, is awe-inspiring beyond all else in nature.' ([7], p. 6)

Bush sees the mechanical, 'artificial indexing' as more organic, more in line with human associative memory. Bush does not want to explain an 'association' in terms of a 'trail', his aim is to transfer the organic and the mechanical characteristics of associations and trails from metaphor into a mechanical device. So he moves back to a mechanical idea (or the idea of mechanization, more precisely said):

'Selection by association, rather than indexing, may yet be mechanised. One cannot hope thus to equal the speed and flexibility with which the mind follows an associative trail, but it should be possible to beat the mind decisively in regard to the permanence and clarity of the items resurrected from storage.' ([7], p. 6)

Bush finds machine-transferable qualities in associations. The organic is extended by the mechanism of marking (indexing) associations as marked trails to prevent them fading.

Applying Black's method of interpreting predicative metaphors, we are able to find similar directions of the meaning interaction:

Vannevar Bush: Association is trail			
Primary subject: association	Secondary subject: trail		
Implications	**Implications**	**Pairing**	**Way of pairing**
association is a connection of thoughts	trail is connection of places	[connection of thoughts, connection of places]	analogy
association is called into mind by symbols, which are given by some convention	trail is equipped with marks	[association's symbol, trail marking]	metaphorical coupling
it is hard to remember associated items without remembering the convention, i. e. by mnemonic devices	it is hard to follow a trail without maps and marks	[mnemonic devices, maps and marks]	metaphorical coupling
a not followed association is prone to fade	a not used trail fades	[association's fading, trail's fading]	analogy

Table 2. Interpretation of Vannevar Bush's mechanic metaphor 'Association is trail'

---

[6] Quoted from Blumenberg [6], p. 67.

The first pair of implications is an example of an analogy. The implication 'trail is a connection of places' is analogous to 'an association is a connection of thoughts'. The primary, organic subject is seen in light of the secondary, mechanical subject. The better-known concept of the trail is extended so that it covers an association. The second pair is an example of metaphorical coupling. Trails are usually provided with marks. Such marks are metaphors for symbols by which associations are called into mind. The third pair may be a case of a metaphorical coupling again. We use marks, or more generally maps, in order to follow trails. In our metaphor we use mnemonic devices in order to follow our association, or to remember associated items. The fourth pair seems to be the case of a metaphorical coupling too: Disused trails fade. This is analogous to a not followed association. They are prone to fade.

Black says, the literal meaning of 'an association' interacts with the literal meaning of a 'trail' resulting in a metaphorical meaning of a 'trail' which is hereby being predicated by an 'association'. The very same metaphor says something about the secondary subject: Bush sees a 'trail' in the light of an 'association'.

'An association is a trail' is a case of absolute metaphors in Blumenberg's sense. It is the unifying representation, which help us to orient in the evolving concept of hypertext. In this stage of discovering hypertext, it is not possible to translate its idea into unambiguous, scientific language. There is no existing technology allowing us to run the memex. There is no scientific terminology yet and it would not be fruitful to establish it. The inventor is only able to show the first orientation of his ideas. In the next step he develops his metaphor into a theoretical model of the hypertextual linking of information, a memex. Nevertheless, a detailed analysis of the memex is a theme for a more detailed investigation. We can only confirm the memex as a fruitful theoretical model in this paper due to the following reasons: The memex resembles the use of metaphors in requiring an analogical transfer of vocabulary. Bush wants to mechanise an organic association trail in his memex. His aim is to improve an organic, transitory memory by means of a mechanical, permanent trail of an association. Bush speaks about the mechanical memex using the terminology of an organic, associative memory. In stretching the language by which the associative memory is described, in such a way, as to fit the new domain (memex), Bush pins his hopes upon the existence of a common structure in both fields. His hope is fulfilled, so there is objective ground for the analogical transfer.

We can describe the memex in the terms of Black's model as a 'system of imaginaries' ([2], p. 234). Bush concentrates on the principle of indexing associative trails. The memex allows the establishing, marking and following of associative trails to be permanent. The memex is supposed to add the organic factors of speed and convenience to the ordinary mechanical filing-system processes. Bush is aware that it cannot work at the same speed as an organic, human memory. But he believes it will be possible in the future, that new technologies will allow future machines to work at the same speed as humans can think. This example shows that Bush was not limited by considering only the real means that were available to him. He built a model, a system of the possible, system of imaginaries. We can consider such a system, pragmatically built as an equilibrium to be consisting of both the organic and mechanical qualities of a human and a machine.

According to Black ([2]), we appreciate the memex as a very vital model. The memex is based on implications rich enough to suggest novel hypotheses and speculations in the primary field of investigation. It suggests further questions, it takes us beyond the phenomena from which we began, and it tempts us to formulate hypotheses which turn out to be experimentally fertile in the future of hypertext development. Bush supposes that clever usage of an associative trail manipulation can augment human associative memory. As we will show in the next section, his concept of associative linking content was inspirational in the questions of human intellect augmentation, by means of a technological extension.

In this section we have analysed Bush's metaphorical thinking in detail. Seen in the broader context of hypertext inventing, the mechanisation of organic qualities of a human mind is essential for contemplating hypertext. In the following section, we will show how the direction of interaction changes. The new direction will lead us to the following question: how can a system of mechanised associations become more organic by means of human machine interaction and cooperation? Will this be fruitful to think about mechanical devices in terms of a text?

5 NLS: INTERACTION BETWEEN HUMAN AND MACHINE

We tried to find some innovative metaphors about content linking for our analysis of Engelbart's text. Nevertheless, Engelbart uses Bush's metaphor mentioned above. In this section we analyse and interpret the metaphor 'a computer is a clerk'[7], abstracted from Engelbart's text *Augmenting Human Intellect: a Conceptual Framework* [8]. We believe that it is helpful in our understanding of the next metaphorics turn and also in the context of hypertext development. We will complete our analysis with an interpretation of Engelbart's NLS system. As will become evident, Engelbart speaks about this machine in the same way as a text, which is an essential direction for hypertext development.

The 'a computer is a clerk' metaphor fulfils our criteria of an absolute metaphor, because it is created as an analogy of an organic and mechanical subject. Allegedly, a computer seems to connote mechanic qualities whereas a clerk is organic, connoted with human qualities. Based on the analysis following Black's interaction theory, we argue that Engelbart turns to see a machine being more organic: as a human being and, in the case of the NLS system, as a text.

Engelbart begins his paper with the task of augmenting the human capability to solve problems:

'By "augmenting human intellect" we mean increasing the capability of a man to approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems.' ([8], p. 1)

Engelbart's main aim is to invent a means that would make the individuals, intellectually more effective, by means of a human-computer interaction:

---

[7] 'Let us consider an augmented architect at work. He sits at a working station […]; this is his working surface, and it is controlled by a computer (his "clerk") with which he can communicate by means of a small keyboard and various other devices.' ([8], p. 70)

'We see the quickest gains emerging from (1) giving the human minute-by-minute services of a digital computer [...], and (2) developing new methods of thinking and working that allow the human to capitalize upon the computer's help. By this same strategy, we recommend that an initial research effort develop a prototype system of this sort aimed at increasing human effectiveness in the task of computer programming.' ([8], p. 3)

Engelbart uses the analogy of a computer as a clerk, as a 'fast and agile helper'[8].

Douglas Engelbart: Computer is clerk			
Primary subject: computer	Secondary subject: clerk		
Implications	**Implications**	**Pairing**	**Way of pairing**
computers have users	clerks have supervisors	[user, supervisor]	analogy
computer is a fast and agile helper	clerk is an agile helper	[helper, helper]	metaphorical coupling
computer is programmed	clerk have to follow rules and laws	[following programs, following rules]	analogy
computers work mechanically	clerks do a lot of mechanical routines	[mechanic work, mechanic routines]	analogy
computers are without emotions and errors	clerks have to avoid emotions and errors	[mechanic, suppressing organic qualities]	analogy

Table 3. Interpretation of Douglas Engelbart's organic metaphor 'Computer is a clerk'

The first pair of implications suggests that computer users are analogous to clerks' supervisors. Engelbart imagines the computer of the future in terms of human collaboration, as a mechanic helper, which needs to be programmed and led by his organic supervisor. The idea of programming is essential in the concept of interaction. The second pair of implications shows that, for Engelbart, a computer is a fast and an agile helper. A clerk is also seen usually as an agile helper. Something mechanical (computer) is analogous to something organic the (clerk). Only mechanical features of clerks are transferred according to this metaphor. We select only the mechanical features of an organic secondary subject. This is going to be explicit in our third implication: Computers work mechanically whereas clerks perform a lot of mechanical routines. The direction of interaction (from 'A to B' or 'B to A') is evident in the last implication. Clerks should be free of emotions in order to avoid errors. They have to suppress their organic qualities and work mechanically. Their mechanised, programmed way of working is now transferred into computers.

Seen from a metaphorological perspective, Engelbart follows his contemporary influential thinkers. Licklider [16] speaks of 'man-computer symbiosis' and Ulam [17] uses the term 'synergesis'. Most comprehensive is Ramo's [18] term 'synnoetics', applicable generally to a cooperative interaction of people, mechanisms and automata into a system whose mental power is greater than that of its components. We find these organic and

mechanical metaphorics to be leading at the beginning of the digital age. Engelbart's text reflects the difficulties with describing his images about the future and possible reality, in the way of literal and scientific terms, Reading between the lines here, he creates his vision in the figurative way of imaginations and he supposes this way to be more comprehensible to his readers.

'The picture of how one can view the possibilities for a systematic approach to increasing human intellectual effectiveness, as put forth in Section II in the sober and general terms of an initial basic analysis, does not seem to convey all of the richness and promise that was stimulated by the development of that picture. Consequently, Section III is intended to present some definite images that illustrate meaningful possibilities derivable from the conceptual framework presented in Section II. The style of Section III seems to make for easier reading. [...] Section III will provide a context within which the reader can go back and finish Section II with less effort.' ([8], p. 3)

However, let us return to the pragmatic reasons for hypertext discoveries. We have to mention Engelbart's account of linking. In the third section of his *Augmentation*, Engelbart comments on Bush's main ideas about a hypertextual content linking, derived from the 'association is a trail' metaphor. From a technical point of view, Engelbart continues in Bush's effort to mechanise linking information by indexing. He broadens this task, because he thinks about links and connections as about interactions. The literal meaning of interactions stresses the meaning of a two-way connection and communication, just like the meaning of feedback. Engelbart with his team was capable of creating a functional, collaborative knowledge environment system called the NLS (for oNLine System). (It was first demonstrated in 1968.) Engelbart's lab used NLS for all its own knowledge work, drafting, publishing, shared screen collaborative viewing and editing, document cataloguing, project management including a shared address book – all of these in an integrated hyper groupware environment. It was possible to edit the structure as well as the text.

While Bush saw the memex as a tangible, a mechanised, a personal library, Engelbart considered the NLS to be an editable text with rewritable links. He saw it as a sort of self-organizing retrieval system, which dealt with the symbolic structures by means of programming.

How does the direction of the organic and mechanical metaphorics interaction change with Engelbart? Engelbart sees mechanical devices in the light of organic, human qualities, interacting by means of symbolic communication. He tries to put the mechanic implications nested in organic terms (i.e. systematization, logic, routines) into machines and augment them. He suppresses (for his pragmatically determined aim) any undesirable organic characteristics in his machine, (i.e. a high error rate, forgetfulness, tiredness etc.). In the next step Engelbart tries to improve mechanical devices by means of suitable organic qualities (i.e. the ability of symbolic communication, ability of feedback, speed of associative processes etc.). In contrast to prior historical eras, he started to explain organic qualities as nested in mechanical metaphorics. Or we can say, the metaphorics of the mechanical is replaced by the metaphorics of programming.

With these thinkers considering pursuing this direction, the metaphor of the mechanical is now becoming corrected (or furthermore developed) by the metaphorics of the algorithmisation. In the next section, we will follow how the text becomes

---

[8] 'Such a fast and agile helper as a computer can run around between a number of masters and seldom keep any of them waiting [...] ([8], p. 70)

hypertextual in Nelson's thinking, and the figurative conceptualising of the new information media.

6 XANADU: ORGANIC MACHINE AS MORTAL MACHINE

In this section we analyse and interpret the metaphor 'a hypertext is a Xanadu', abstracted from Nelson's hypertextual project [20]. Nelson coined the term 'hypertext' and defines its properties in 1965 ([21], p. 96). In *Literary Machines* ([9], p. 30) he describes his most famous hypertext project Xanadu as a 'magic place of literary memory'. His hypertext concept is supposed to be analogical to Coleridge's Xanadu [22]. We will concentrate on Nelson's implications from this metaphor.

Nelson wants to transcend the possibilities of textual form, determined by the qualities of mechanical printing machines. The metaphor, which he chooses, answers this purpose. We can see the connection with Engelbart's approach. Nelson and he sees a machine as a text. While Engelbart only notices this analogy, Nelson is able to develop it in a very detailed way with the help of figurative language, but also in unambiguous, scientific definitions of hypertext qualities. The word 'hypertext' we can consider as specific type of metaphor, catachresis, which, according to Black, fulfils the gap in the existing vocabulary. As with Musil's man from the epigraph, with a sense for the possible, he abstracts from the given (mechanical) reality which is insufficient for him:

'The sense of "hyper-" connotes extension and generality; […] The criterion for this prefix is the inability of these objects to be comprised sensibly into linear media […].' ([21], p. 98)

Hypertext is the presentation of information as a linked network of nodes which readers are free to navigate in a non-linear (organic, associative, creative) fashion. Nelson does not want to mechanise the organic, as Bush did. Most of all, he wants to create a new, more organic, more human media. Which organic qualities does he transfer into his literary machine, i.e. hypertext? He wants to teach machines human skills such as writing and reading. The Xanadu user is the reader and the writer of the text at first. And he is a programmer too. As Fuller and Goffey [23] show, programming is a new use of a language and the language has a very organic, human quality.

Ted Nelson: Hypertext is Xanadu			
Primary subject: hypertext	Secondary subject: Xanadu		
Implications	Implications	Pairing	Way of pairing
hypertext concept is rich	Xanadu offers a lot	[rich, offers a lot]	analogy
hypertext is a text with a new dimension	Xanadu is a magic place	[new dimension, magic place]	metaphorical coupling
hypertext is a text with references to other texts	Xanadu is a place of literary memory	[web of texts, literary memory]	analogy

Table 4. Interpretation of Ted Nelson's organic metaphor 'Hypertext is a Xanadu'

Nelson explains his hypertext as a Xanadu. The first pair of implications suggests that the concept of hypertext is as rich as a Xanadu. The second pair of implication-complexes is a case of

metaphorical coupling: a Xanadu is a magic place in Coleridge's poem, while Nelson's hypertextual Xanadu adds a new dimension to the text. Coleridge's Xanadu transcends the materiality of our world, hypertext remediates materiality of 'paper' with its qualities. The third pair of implications defines Xanadu as a place of literary memory. This is analogous to hypertext being a text with references to other texts. Coleridge's Xanadu is a metaphor for the never-ending finding of a magical place. It is dedicated to active and creative users. It functions, after forty years of development in a limited version. It will stay in a dream as in Coleridge's Xanadu. It is too difficult to be the main principle of the contemporary leading hypertextual system, the more mechanical WWW. As Nelson says:

'Today's popular software simulates paper. The World Wide Web (another imitation of paper) trivializes our original hypertext model with one-way ever-breaking links and no management of version or contents.' ([20])

The reason is pragmatic: for general purposes we need an easier solution. In this aspect, the historical dialectical interplay of metaphorics, at the turn of the twentieth and the twenty-first century, shows us that a more mechanical medium is more vital than an organic one. But Xanadu has a chance to inspire a specialised, professional system for scientists and people who have to think in a more complex way. Or, we can change the direction of metaphorics, and go along with Rushkoff, to suit people, who do not want to be programmed, but want to programme [24].

7 CONCLUSIONS & FUTURE WORK

The common pattern of the analysed metaphors in Black's interactive view is that the interaction of the meanings in them goes in two ways. The implications of the mechanical and the organic metaphorics are nested one in the other and therefore these metaphors are absolute in the Blumenberg sense. The interpretations of our metaphors are open-ended and fruitful for new concepts of hypertext. We applied this idea in models and concepts of hypertext: All of our hypertext thinkers speak about the human-machine interaction in terms of finding the best equilibrium of the possible and the real, of organic and mechanical qualities. The direction of their investigations leads from the need of a mechanical machine, based on organic principle to a new medium, based on the transfer of many human organic qualities and skills into an interactive medium.

Bush mechanised the way of human, organic associative indexing and makes mechanical ways of indexing more organic, more in line with human thinking. On the other hand, he contemplates the mechanisation of associations.

Engelbart's hypertextual equilibrium stresses the interaction of human (organic) and mechanical (computerized) elements. He speaks mostly in terms of mechanic qualities nested in organic, human elements. He stresses the idea of seeing a machine as an (organic) text, as a medium.

Nelson builds upon his predecessors' idea, that the medium is more organic. He wants to transcend the possibilities of the textual form determined by the qualities of mechanical printing machines. He speaks about hypertext in more organic terms. His concept is very organic and therefore mortal, as we have shown.

In the period in question, the history of the concept of hypertext started with an organic metaphor of association. It continued through the idea of mechanisation and furthermore through the

idea of organic-mechanical interaction and was complemented by the organic metaphorics of reading, writing and programming. In the context of hypertext discovering, a mechanical solution became insufficient. This insufficiency is supposed to be filled by an organic solution. The next step consists in the mechanization of organic qualities, and the following one in their algoritmisation in the era of digital media.

Absolute metaphors, as metaphors in general, fulfil the function of stressing some aspect of the source domain. This function is pragmatically determined. In our case the pragmatic reasons are the following:

(1) to augment human intellect by mechanical means,

(2) to enable other people to understand such difficult thoughts, as Musil's 'unawakened realities', which are not translatable into the literal language of science. ([1], p. 12)

The history of media is the history of attempts at understanding human, organic qualities and to use them as extensions by transferring them into machines. After a successful transfer, the direction of this interaction then changes. Now we start to use media as a translation, as a metaphor for explaining human, organic qualities. It seems that in the era of algoritmisation[9], the metaphorics of mechanical machines have lost its importance. It has been corrected by the metaphorics of the digital media, just as the metaphorics of linear (mechanical) text has been corrected by the metaphorics of (organic) hypertext.

In our future work we will continue pursuing the history of this metaphorics in relation to the WWW. We expect to interpret it as a mechanised organic medium of Nelsonian hypertext. We see the importance in investigating more unique hypertexts such as scientific ontologies.

REFERENCES

[1] R. Musil. *The Man without Qualities*. Trans. by Sophie Wilkins. Vol. I. Vintage Books, London, UK (1996).

[2] M. Black. *Models and Metaphors: Studies in Language and Philosophy*. Cornell University Press, Ithaca, USA, pp. 219–243 (1962).

[3] T.S. Kuhn. Metaphor In Science. In: *Metaphor and Thought*. A. Ortony, (Ed.). Cambridge University Press, Cambridge, UK, pp. 409–419 (1979).

[4] R. Boyd. Metaphor and Theory Change. In: *Metaphor and Thought*. A. Ortony, (Ed.). Cambridge University Press, Cambridge, UK, pp. 481–532 (1979).

[5] H. Arendt. Metaphor and the Ineffable: Illumination on 'the Nobility of Sight'. In: *Organism, Medicine, and Metaphysics*. S.F. Spiecker, (Ed.) Springer, Dordrecht, Netherlands, pp. 303–316 (1978).

[6] H. Blumenberg. *Paradigms for a Metaphorology*, trans. R. Savage. Cornell University Press, Ithaca, USA (2010). For orig. see: *Paradigmen zu einer Metaphorologie*. Insel, Frankfurt am Main, Germany (1999).

[7] V. Bush. As We May Think. *Athlantic Monthly*, July 1945, 176 (1):101–108 (1945).

[8] D.C. Engelbart. *Augmentation Human Intellect: A Conceptual Framework*. SRI, Menlo Park, USA (1962).

[9] T.H. Nelson. *Literary Machines*. Mindful Press, Sausalito, USA (1981).

[10] V. Bush. Memex Revisited. In: J.M. Nyce and P. Kahn (Eds.): *From Memex to Hypertext*. Academic Press Professional, Inc. San Diego, CA, USA, pp. 197–216 (1967).

[11] M. Black. More about Metaphor. In: *Metaphor and Thought*. A. Ortony, (Ed.). Cambridge University Press, Cambridge, UK, pp. 19–43 (1979).

[12] G. Lakoff and M. Johnson. *Metaphors We Live by*. The University of Chicago Press, Chicago – London, USA – UK (1980).

[13] T. Hobbes. *Leviathan*. Critical edition by Noel Malcolm. Oxford University Press, UK (2012).

[14] S.E. Toulmin. *The Philosophy of Science*. The Mayflower Press, London, UK, https://archive.org/stream/philosophyofscie032167mbp#page/n43/mode/2up. (1953).

[15] L.B. Alberti. *On the Art of Building*, trans. J Rykwert, N. Leach and R.Tavernor. MIT Press, Cambridge, USA (1994).

[16] J.C.R Licklider. Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, (March 1960).

[17] S.M. Ulam. A *Collection of Mathematical Problems*. Interscience Publishers, Inc., New York, USA (1960).

[18] S. Ramo. The Scientific Extension of the Human Intellect. *Computers and Automation*, February, (1961).

[19] J.M. Lawler. *Metaphors we compute by*. http://www. ling. lsa. umich. edu/jlawler. (1987).

[20] T.H. Nelson. *Project Xanadu*. http://xanadu.com. (2014).

[21] T.H. Nelson. *A File Structure for The Complex, The Changing and the Indeterminate*, Proceeding of the ACM 20th National Conference, ACM, August. pp. 84–100 (1965).

[22] S.T. Coleridge. *Christabel: Kubla Khan, a Vision; The Pains of Sleep*. William Bulmer, London, UK (1816).

[23] M. Fuller and A. Goffey. *Evil Media*. MIT Press Cambridge, USA (2012).

[24] D. Rushkkoff. *Program or Be Programmed*. OR Books, New York, USA (2010).

---

[9] According to Fuller and Goffey, it is a process of a reality that occurs in the conversion of process, which we know from the physical world into sequence of writable and readable algorithms that drive our human-media interaction. ([23], p. 80)

Metaphor, Meaning, Computers and Consciousness

Stephen McGregor[1] and **Matthew Purver**[2] and **Geraint Wiggins**[3]

Abstract. This paper seeks to situate the computational modelling of metaphor within the context of questions about the relationship between the meaning and use of language. The results of this pragmatic assessment are used as the theoretical basis for a proposed computational implementation that seeks metaphor in the geometry of a vector space model of distributional semantics. This statistical approach to the analysis and generation of metaphor is taken as a platform for a consideration of the fraught relationship between computational models of cognitive processes and the study of consciousness.

1   Introduction

Aristotle is commonly credited as the earliest thinker to seriously consider metaphor as a linguistic device, lauding its use as an indication of the highest level of genius [1]. But while a historical account of scholarship about metaphor is a worthwhile topic, and one which will feature throughout this paper, the history of metaphor itself is as convoluted and unobtainable as the history of language. In fact, if it serves any purpose to think about such a remote event as the inception of language, it seems impossible to imagine a clever speaker not immediately taking the agreed definitions of the world's first words and doing something unexpected with them. If anything, a more accurate take on early academic discussion of metaphor might be to consider Aristotle as one of the first philosophers to ponder the question of the relationship between what words mean and what words do.

This paper will seek to evaluate metaphor from a pragmatic point of view, and to situate this evaluation in terms of a framework for the computational analysis and generation of metaphor. This marks a shift from what has become the standard computational approach to metaphor, which considers language in terms of formalisms that are intuitively compatible with symbol manipulating machines. Implicit in these standard approaches is the assumption that words and concepts exist on different levels of abstraction, and that metaphor is a product of a process of transference or mapping that occurs on the conceptual level, with words acting as a kind of index of this process. But the idea that words merely point to concepts runs into trouble in light of certain properties of metaphor that cannot be explained in terms of an abstract conceptual construct of the entities nominated by words. At the root of the approach proposed in this paper is a contention regarding the difficult topic of consciousness: metaphor is often based on the direct experience of perception, and the ease with which a cognitive agent can express the actual quality of one particular percept in terms of the idea of another general percept is rooted in the direct connection between phenomenology and language. The very relevance of the term "like" to figurative language, manifest when metaphor is translated into simile, suggests that the "likeness" of the conscious experience of qualia is intrinsic in the perpetually unfolding construction of metaphor.

One of the claims made in this paper is that consciousness is always understood metaphorically, and one of the most pervasive and at the same time disputed contemporary metaphors involving consciousness has been the trope that casts the mind as a computer. This particular construct is compelling, in that the mind can be conceived of as having input in the form of perceptual stimulation and output in terms of either conceptualisation of the world or directed action in the world. At the same time, the analogy is disreputable in its relegation of the richness of consciousness to the domain of a rule following, data processing apparatus that is subject to an arbitrary, observer relative interpretation. It seems that a good model of metaphor should explain the appeal of comparing the engine of its own operation – the mind – to a device that is arguably at best just an aid to thought. The model should also account for the perceptual, imagistic aspect of metaphor-making, evident in light of the necessity of comparing one experience to another when trying to describe what it is like to be conscious.

The solution offered here involves turning to high dimensional representations of meaning based on a statistical analysis of the distribution of words in large scale corpora, and, in so doing, embracing the modelling power of the computer, if not the explanatory power of the mind-computer metaphor. The theory behind the system that will be described is based on the idea that a statistical treatment of a large collection of words found in their natural habitat, so to speak, can simulate the construction of a space of meanings. This space, in turn, becomes the linguistic environment in which metaphors are discovered in the process of solving communicative problems: congruences in the geometries of these statistical word-objects suggest ways in which they can be combined in order to construct expressions. The metaphor-making procedure, modelled as a fundamental aspect of ongoing entanglement with a richly informative environment, is finally presented as a key component in the expression of consciousness, a characteristic that may shed some light on the evident propensity for qualia sensing agents to project their own consciousness onto everything else in the world.

2   Consciousness Is a Metaphor

The tension that metaphor has traditionally introduced to the study of language has arisen from the dynamic between words and truth: figurative statements that are clearly contrary to the facts of reality are nonetheless effective at conveying truthful information about the world. This aspect of metaphor poses at least a superficial problem for truth conditional approaches to semantics, which hold that there is either a correspondence between propositions and the world they

---

[1] Queen Mary University of London, email: s.e.mcgregor@qmul.ac.uk
[2] Queen Mary University of London, email: m.purver@qmul.ac.uk
[3] Queen Mary University of London, email: geraint.wiggins@qmul.ac.uk

portend to describe [36], or a coherence between the set of propositions that collectively constitute a truthful system of beliefs [13]. For Floridi, the imperative of truthfulness means that "semantic information" is necessarily defined in terms of data that remits "veridicality" in relation to the world that it models [19]. Dretske likewise distinguishes between information and the the semantic representation indicated by a correct interpretation of that information [16].

Taking Dretske's ideas about indication and interpretation as a point of departure, it is possible to formulate a theory whereby the truthfulness of figurative propositions lies in the correct interpretation of the intention behind a non-veridical statement. Here metaphor becomes a mechanism for encoding information, with the projection from source to target allowing for the transference of a set of intensions from a general case of the source to a specific instance of the target. If this is the case, then a metaphor can be deciphered into a more extensive array of literal propositions. The well studied metaphor "that surgeon is a butcher", for instance, takes the bloodiness and brutality stereotypically associated with the profession of a butcher and efficiently applies them to the behaviour of some disreputable surgeon. This packaging of literal information sits well with Searle's approach to metaphor, which sees non-literal language as an invitation to interpretation based on propositional knowledge of the world shared between two interlocutors [35]. Ortony, in his "reconstructivist" theory of metaphor, has even suggested that there must be some sort of mental imagery involved in the interpretation of figurative language: a metaphor evokes a non-literal scene which effects the vivid transference of intension in a way that invites logical inference [31]. This move introduces conscious perception to the explication of metaphor, with the experience of a mental state playing a direct role in the transmission of richly detailed information.

But how can consciousness ever be discussed in a way that is literal or veridical? If qualia, with their intrinsically subjective character, are the substance of conscious experiences, then it seems impossible to describe such phenomenological conditions in terms of truthful propositions about situations in the world. Chalmers has made much of this divide between subjective conscious experience and objective physical reality, focusing in particular on the difficulty of determining the truth conditions of a report of a phenomenological perception [9]. From a phenomenological perspective, the defining characteristic of consciousness is that there is something it is like to experience qualia, and this very "likeness" of the experience immediately suggests the application of analogical conceptualisation and correspondingly metaphoric expression. While a mutually agreed description such as "red thing" might allow two interlocutors to pick out a set of objects with some shared characteristic, it is not clear that there is any way to know that the actual phenomenology of the red experience is similarly shared. Since there is no way to expressively project the actual conscious experience of perceiving an object, a descriptive speaker who wishes to convey something phenomenological is left with no choice but to resort to an act of analogy, giving the world such poetic turns of phrase as "lips as red as blood" or "eyes as blue as the sky".

Along these lines, Everett has highlighted the absence of abstractly quantifiable colour terms in the language used by the Pirahã people of Brazil, who instead employ standardised expressions that are fundamentally figurative: the color term corresponding to what an English speaker would describe as "red", for instance, transliterates to the expression "bloodlike", and "black" becomes the phrase "blood is dirty" [17]. Levinson reports similar findings in his analysis of the Yélî Dnye language spoken by the inhabitants of an isolated island near Australia, who use the terms for various birds and plants

to describe other similarly coloured objects [29]. Even if, as Kay and Maffi claim, the lack of fixed absolute colour partitions in a language is anomalous [27], the admission of chromatic descriptions such as "chartreuse", "coral", or "eggplant" in English illustrates the ease with which a perceptual experience of one thing can be converted into a classification of something else. There is an inherent process of analogising occurring when cognitive agents turn to language to express the subjective characteristics of their perceptual existence.

This perpetual trafficking of intension from one perceptual or conceptual domain to another extends especially into more general descriptions of consciousness. The difficulty of discussing qualia in objective and material terms has compelled philosophers to resort time and again to thought experiments involving components fantastically removed from reality – beetles in boxes, homunculi in theatres, deceptive demons – in order to allude circumspectly to what it is like to be conscious. Even Dennett, who has questioned the efficacy and indeed the existence of qualia [14], acknowledges that it is generally necessary to employ analogical reasoning when dealing with descriptions of mental processes [15].

There is a temptation to take the necessity of analogy in discussion of consciousness one step further by way of construing consciousness itself as a process of metaphor-making. In the 1970s, Jaynes proposed his bold "bicameral" theory of mind based on the idea that pre-literate humans had perceived their own consciousness as a mentally external expression of instructions and proclamations experienced as ongoing auditory hallucinations [26]. To a mind sundered in such a way, the modern experience of self as realised through subjective phenomenology was supposedly replaced with a personal fictive narrative that cast the consciously feeling component of the mind in the role of a god or a commanding spirit. This controversial theory has received some recent support, at least implicitly, in Carruthers' formulation of "interpretive sensory-access" based on the mindreading faculties that facilitate the acts of interpretation at the centre of consciousness [6]. In a propositional reversal that nonetheless maintains some of the core tenets of Jaynes' bicameral mind, mindreading capacities can be applied not only to introspection, but also to the interpretation of the mental states of other people and even as the projection of mind-like faculties on objects that are obviously actually inanimate. So, for instance, it seems quite reasonable to metaphorically discuss the temperament of things like computers, cars, appliances, or the weather without the presumption that these types of objects actually have minds.

If these projective theories of mind are to be taken seriously, then the essential role of metaphor in consciousness must be considered. There certainly seems to be a case to be made for the idea that consciousness necessarily involves a transgression of literal conceptualisation of the world, a transference of a feature from one mental object to another that results in an expression of the experience of a thing as something other than what it actually is. There are three propositions at stake here. The first is that the only feasible mechanism for communicating about the experience of consciousness is to cast the description of that experience out onto some universally accessible entity with qualitative attributes that will hopefully simulate the experience. The second is that the mind can only be understood in terms of things other than minds, things that have mind-like properties and therefore analogically corroborate an explanation of what it is like to have a mind. The third is that having a conscious mind necessarily involves the projection of phenomenological characteristics onto external entities, some that presumably are likewise conscious and others that almost certainly are not. In each of these cases, through experiential transference, through analogical descrip-

tion, and through projection of the self onto another, an essentially metaphoric process is at play: knowledge of the mind seems to consist of a network of proxies and equivalences that trace the outline of the thing that they don't quite touch.

3 Words Are Objects

The recent history of theoretical approaches to metaphor has been characterised by an intellectually productive tension, with both sides notably departing from any notion that the figurative use of language should somehow be treated as an exceptional case. On the one hand, there are those who would describe metaphor as a transference or projection of intensionality from the conceptual space of a source to the similarly oriented space of a target, a view that found an early champion in Black and his "interactionist" theory of metaphor [3, 4]. By this account, metaphor involves conceptual mappings that place a non-literal source at the centre of the "implicative complex" of a targeted conceptual system, so that characteristics of the way the source does things are projected onto similar activities undertaken by the target. On the other hand, a dissenting contingent of theorists have argued that the metaphoric use of language stands entirely outside the realm of conceptualisation, and that the meaning of any sentence can only be interpreted literally—an idea originally expounded by Davidson [12], with early support coming from Rorty [34].

In the early 1960s, Hesse argued for the importance of analogy as a tool for scientific understanding [25]. At the root of her argument was the idea that all theories are ultimately models of the world, and that, in terms of the extreme scales involved in, for instance, the study of physics, these models could only be grasped in terms of metaphors: so, for instance, a distributed gas bears an analogy with a space full of colliding and rebounding balls. The study of metaphor subsequently underwent a Renaissance of sorts, with a flurry of research throughout the 1970s (see [32] for a compendium of exemplars), culminating in Lakoff and Johnson's case for an understanding of metaphor as a mapping between isomorphic conceptual schemes [28]. This theory presented metaphorical language in terms of its relationship to an embodied cognitive experience of the world, so, for instance, the analogy which maps the conceptual situation between "up" and "down" to the situation between "happy" and "sad" is a product of the actual culturally loaded experience of orientation in the real world. A lattice of networked spaces, extending from the world through perception and conceptualisation into language, allowing for the transference of entire isomorphic conceptual complexes: if a surgeon is a butcher, then hospitals become abattoirs and patients become animals.

Davidson, however, offered a dissenting interpretation of metaphor, springing from his rejection of the idea that language should be talked about as a system for conceptual representation in the first place [11]. Instead, he proposed that the meaning of a metaphor could only be considered in terms of the literal proposition made by a metaphorical statement, and that the operation of a metaphor in the process of communication must be considered as something altogether outside the realm of meaning [12]. This stance has met with considerable resistance, finding an early opponent in Bergmann, who argued that Davidson's critique only applied to decontextualised encounters with metaphor; once the metaphor is put into the context of a situation involving a speaker with an intention, it can be clearly seen to have a meaning [2]. Hesse also revisited her case for metaphor as a fundamental cognitive operation, arguing that all language is metaphoric in that all language plays a protean role in a nebulous network of meaning [24]. Rorty, on the other hand, came

to Davidson's defence, interpreting his approach as placing metaphor actually in the world of natural events rather than consigning it to an essential role in an interplay of symbols that is ancillary to reality [34]. By this reading, language is not to be considered as a model or representation of reality, but rather as a component directly in reality, existing on the same level of abstraction as impressions and ideas.

The debate over metaphor in subsequent years has involved a back and forth between those who see metaphor as by-product of an essential cognitive operation and those who claim that language plays a more fundamental role in perception of the world, though Davidson has arguably been broadly misinterpreted. In an expansive consideration of metaphor as evidence of "the poetic structure of mind", Gibbs suggests that Davidson places emphasis on first determining the literal meaning of a metaphor and then accepting that the potential non-literal meanings of the phrase are somehow infinite and unknowable [22], perhaps a misreading of Davidson's contention that "there are no unsuccessful metaphors". As a recent proponent of the non-cognitive take on metaphor, though, Carston has recast Davidson's rejection of cognitive content in terms of a more fundamental "imagistic" feature of language [7]. In particular Carston considers the metaphor "Bill is a bulldozer": the interpretation of this phrase as a description of a man who is grossly aggressive and inconsiderate is clear, but upon further analysis there is no literal property of a piece of equipment such as a bulldozer that bears the inherently human intensions being drawn out in Bill [8]. At best there might be an argument that a double metaphor is being employed here, with a bulldozer standing in for something aggressive and then Bill being described as one of those things, but this introduces a combinatorial explosion of ways to frame all but the simplest metaphors and in so doing seems to miss the point of the cogency of figurative language. Instead, it seems reasonable to say that the metaphor evokes something that is not purely in the realm of language, a direct perception of Bill as a potentially destructive machine.

In this analysis, Davidson and his acolytes emerge as something of the arch-pragmatists. Rather than keeping the construction and interpretation of metaphor on a symbolic level, where language models the world it describes, here the very meanings of the words employed in a metaphor become implements to be handled and used to accomplish communicative goals in the same *ad hoc* way that a more overtly physical object might be picked up and used. Meanings exist, but as the features of elements of language that suggest their functionality: in fact, the meanings of words themselves become the intensions of those words, suggesting potential uses of language in the way that, for instance, the solidness and heaviness of an object might recommend it as a weapon to an attuned perceiver in need of such a device. Just as a shoe might present itself as a hammer under the right circumstances, or a stick or rock as a writing instrument, the word "bulldozer" offers itself as the right term to convey Bill's comportment in the same grasping process of perception and cognition, because language is actually happening on exactly the same level as the rest of existence, not in an abstract secondary space.

At this point, language can be situated in the context of Gibson's theory of affordances, which holds that cognition arises in the process of the perception of opportunities for action in an environment [23]. Clark has worked towards expanding environmentally situated approaches to cognition into the domain of linguistics, describing the "persisting but never stationary material scaffolding" of language [10]. A picture emerges of language use as a process of scavenging a shifting space of meaning for the words that can be used to accomplish some expressive task. These meanings are not representational models that stand in a relationship of signification to perceptions and

conceptions of the world; they are the cognitive detritus of entanglement in an environment that involves communication with other linguistic agents, sitting right alongside other mental experiences of reality.

So an alternative approach to modelling metaphor emerges, one that does not involve considering the language involved in metaphor-making as simply a corollary to mappings between isomorphic conceptual spaces. Instead, metaphor can be envisioned as a process of searching a space of linguistic percepts for the sounds or symbols that can be arranged to fulfil some communicative requirement. The challenge then becomes defining this space of meanings and understanding how word-objects are selected from it. This theory does not refute the descriptive power of Lakoff and Johnson's ideas about conceptual metaphors; in fact, it seems clear that there must be some discernible aspect of meaningful entities that allows them to be cobbled into a pragmatically efficacious structure, and it seems likewise reasonable to construe this act of construction as an aligning of mental objects. As an explanatory device, though, the idea that metaphoric language simply corresponds to congruent concepts seems, upon closer analysis, insufficient.

Hesse's quip about all language being metaphoric also follows from this revised approach: all language use involves grabbing meanings that present themselves as functionally appropriate for the communicative act at hand, and, while some constructions may challenge interpretation more than others, there is no clear reason to draw a definitive line between the literal and the figurative use of meaning. The ubiquity of metaphor takes on a more distinctly Peircean character, though, when word-objects are recognised as existing in the same cognitive space as other percepts. Peirce's claim that all thought is realised through signs [33] seems of a piece with Davidson's pragmatic approach to metaphor once the difference between considering objects as symbols of the mind versus considering symbols as objects of cognition becomes a relatively minor point of contention. To Peirce, reality was a lattice of ubiquitous signification, with meaning manifesting itself through a "life in signs", by which all thought results from the inherently interpretable interplay between things, and all physical interactions are characterised by this kind of life. The perpetual life cycle of event, perception, and interpretation means that signs are always exploding outward from the thing that they signify, becoming themselves the object of a further signification in the instant of their interpretation, even as the interpretation becomes a sign of the thing it interpreted. This endless sequence of becoming something else, accomplished by means of the transformative faculties of symbols, points to a fundamental and enduring process of metaphor-making in the experience of existence.

And here consciousness re-enters the consideration of metaphoric language: consciousness as the thing that can only be objectively grasped through metaphor, or metaphor as the mechanism that facilitates the subjective experience of consciousness. By Peirce's account, the world is conscious, an audacious asseveration that nonetheless lines up well with the idea that being conscious involves the perpetual invocation of the fundamental metaphor that everything else is conscious, as well. If the Peircean variety of panpsychism is perhaps a bit strong, a consideration of the metaphoric nature of individual consciousness at least offers an explanation of why the rest of reality would seem that way, as well. In fact, in accepting that language is wrapped up in a pragmatic process of meaning-grasping, and that all use of word-objects is essentially a ready-to-hand encounter with linguistic percepts, the experience of perpetual metaphor and therefore of imminent and ubiquitous consciousness becomes a less alarming outcome.

4 Meaning Is Geometric

Computational models of metaphor have tended to embrace symbolic approaches that treat language as a representation of cognitive content. As a first approximation, this is not unreasonable, given that computers are symbol manipulating machines: a semantic formalism is precisely the kind of information processing model that is tractable to such a machine. Consequently, van Genabith has found success programming computers to analyse metaphors using type theoretical constructs where source and target both belong to a supertype by virtue of their joint properties, and the intensions transferred by the metaphor are categories specific to the type of the source [38]. Similarly, Veale has built a computational system that handles analogies in terms of "dynamic types" culled from prefabricated conceptual networks such as WordNet [39]. And Gargett and Barnden have described a metaphor generating system that applies information processing instructions to conceptual structures [21], in this case implemented through the contextually sensitive typed schema of Feldman's embodied construction grammar [18].

These kinds of systems treat words as indices to concepts, where the logical structure of concepts can be aligned so as to indicate the affiliated linguistic expression that conveys the projection of properties from source to target. In this way, they are implementations of the conceptual approach to metaphor outlined by Lakoff and Johnson: their success stems from their recourse to abstract representations of concepts, and language is treated as a kind map of the mappings inherent in the dynamics of the conceptual space, metaphoric precisely because of the analogical aspect of cognitive content. In the case of Veale's system, the conceptual schema are, compellingly, built in an *ad hoc* way, even if this ongoing construction is based on a pre-established network. With Gargett and Barnden's system, the underlying formalisms are specifically designed to contextualise conceptual representations in terms of the physical world. By the same token, though, these models are intrinsically committed to the cognitive-content approach to metaphor, treating language as a secondary feature merely pointing to the world model of a conceptual space.

It is not clear how such a system could, for instance, model the direct imagistic experience of perceiving an aggressive person as a bulldozer. The inescapable figurativeness of consciousness, that property by which there is a bulldozer-like quale in the encounter with this unpleasant individual, is lost to a system that depends on conceptual constructs removed from encounters with the percepts – the language and the imagery – that become the symbolic index to those concepts. If the project of computationally modelling metaphor is to be pursued further, it seems necessary to formulate a way in which a space of meanings can be constructed directly from an encounter with language in the world, based on the actual statistical features of the language rather than on predetermined rules regarding the processing of symbols. But how can a computer go about realising this kind of language model?

In fact, symbol manipulating machines seem like exactly the right tools for engaging with this task, and a viable methodology already exists in the form of ongoing work on vector space models of distributional semantics. This approach to language modelling involves the geometric representation of words as points in a high dimensional space [40]. Words are construed as vectors, with the dimensions of these vectors corresponding to the contexts in which a word is likely to occur: in the most straightforward implementation, a dimension of a word-vector corresponds to a term, and the scalar value of that dimension indicates the likelihood of the word co-occurring with

that particular term. When the co-occurrences of the words found throughout a large scale corpus are computed, the result is a space in which the proximity of word-vectors to one another corresponds to the similarity of the contexts in which those words have been found. The intuition behind work in this direction has been that words that are found in a similar context will naturally be likewise semantically similar [37].
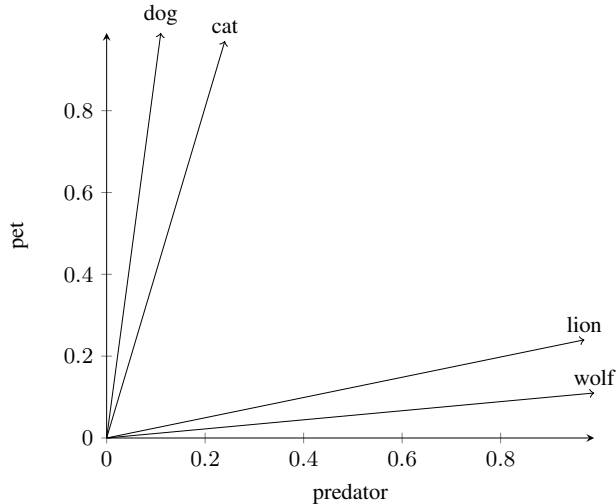

Figure 1: In a highly simplified (two dimensional) vector space model, the words "dog" and "cat" are seen to be semantically relatively similar by virtue of their frequent co-occurrence with the term "pet", whereas "wolf" and "lion" are more likely to occur in the context of the term "predator".

Furthermore, the mathematically tractable properties of a geometric space have been exploited in the modelling of compositionality, with linear algebraic operations between word-vectors producing statistical structures corresponding to the meaning of larger segments of language [5, 30]. A similar intuition can be applied to the construction of metaphor, though with the philosophical caveat, informed by Davidson's take on metaphor, that, where meaning applies to the space of words, the compositions constructed from this space are properly understood only in terms of their use in acts of communication. All the same, it is the geometry of the space of words that suggests ways in which sets of meanings can be pragmatically constructed as metaphors: if proximity corresponds to similarity, then regional clusters of related terms should be discoverable within the vector space. Moreover, the relationship between the terms within such a space indicates a particular geometry, and a congruence in the configuration of terms between two regions might be interpreted as an indication of a potential metaphor. So, for instance, the constellation of word-vectors indicated by the sequence $\{surgeon - patient - hospital - scalpel\}$ would be expected to line up with the shape described by $\{butcher - animal - abattoir - cleaver\}$.

Underwriting this statistical prediction is the theoretical intuition that the way in which a computer encounters symbols in a corpus stands in a synecdochical relationship to the way in which a cognitive agent encounters percepts – including linguistic symbols – in an environment. The hope is that treating large scale corpora as a kind of native habitat for computers serves as a more veridical simulation of the process by which cognitive agents directly grapple with lan-

guage in the physical world than does the construction of abstract conceptual representations. Just as an agent maintains a shifting lexicon of meaning based on a continuous entanglement with language percepts, a computer can establish a network of relationships based on the statistics of its ongoing encounter with symbols in a textual environment. The statistical relationship of words learned by a corpus traversing computer becomes its knowledge base, its space of meanings that can be invoked in a disengaged way when the definition of a particular term is sought, but that at the same time stand ready-to-hand waiting to be grabbed as affordances in the construction of communicatively effective language. When the moment comes for the system to compose an expression, it grasps for the combination of terms that fulfil the required criteria, and these criteria are specifically modelled in terms of the geometric alignment of regions within the space of meanings.

Some preliminary work has been done exploring the relationship between established conceptual metaphors construed in terms of the arrangement of their components within a vector space model, comparing, for instance, the region of butchery to the region of surgery, or the region of orientation (e.g. $\{up - down - in - out\}$) to the region of emotion (correspondingly $\{happy - sad - inclusion - loneliness\}$). Early results have invited cautious optimism: the geometry of the compared vector regions has remitted a high degree of congruence in the anticipated alignments. Future research will have to examine the way in which regions of vectors, corresponding to the construct of conceptual spaces [20], can be defined within a vector space, and this direction of inquiry will in all likelihood motivate a close consideration of the techniques employed in the construction of the vector space itself, as well. The prospective outcome of this project is a system that will use corpus analysis to facilitate a program outputting novel and useful metaphors based on inputs that are perceived as being relatively literal.
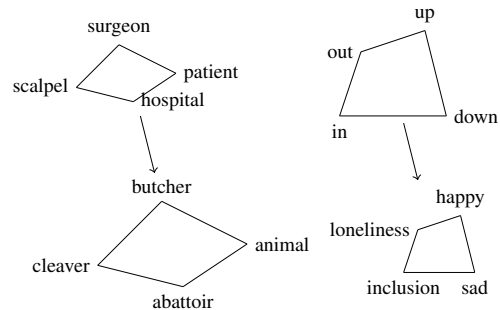

Figure 2: Congruences discovered in subregions of a vector space model suggest metaphoric mappings. The regions do not necessarily have to be of the same scale in order to identify a possible alignment.

5 Conclusion

In weighing the merits of considering the use of words as distinct from the meaning of words, it is worthwhile to observe the extreme ease with which people produce and digest figurative language: metaphor is so universal that almost nothing makes sense if it is taken absolutely at face value. Such a linguistic environment might appear particularly hostile to so formal and literal an agent as a computer. It would seem that the relationship between language and the situations described by language is much messier than some semantic

AISB Convention 2015: 8th AISB Symposium on Computing and Philosophy: The Significance of Metaphor and Other Figurative Modes of Expression and Thought

44

formalisms would suggest, and the role that meaning plays in the process of communication cannot be easily situated in a denotational relationship to some sort of mental content outside of language. In order for a computer to have a chance in a scenario where all language is open to interpretation, it is necessary for the information processing system to have recourse to its own semantic constructs, and these naturally take the form of statistical interpretations of the bearing of words in their compositional contexts.

Using a computer to model the pragmatic dynamics of metaphor reveals nothing about how consciousness works or why consciousness exists. In this regard, the most that can be said about the system described in this paper is that it attempts to simulate a process with which consciousness is concerned—and this much is true of any computer program that presents data in a way that is designed to be interpretable to a conscious user. Nonetheless, the project of constructing a metaphorical framework within a symbol manipulating system takes on added resonance when considered in the scope of the ineluctably analogical modality of the understanding of the conscious mind. Even if the model that has just been proposed doesn't shed any light on the nature of consciousness, it does address some of the questions about the linguistic operation involved in conceptualising consciousness. It is the very ineffability of consciousness that forces a philosopher to resort to analogy and metaphor when discussing this hard topic and indeed when describing the experience of it. In understanding the construction of metaphor as a utilisation of meaning towards the goal of expression, it becomes clear how a cognitive agent must be constantly involved in this operation, always grasping for the combination of meanings that work when put out into the world as the communication of a mental experience. In the process of constructing the sustained sense of self at the core of a conscious experience of the world, a cognitive agent must necessarily cast the idea of the self out into the world to reflect upon it; it is only natural, then, that an essential feature of consciousness should be to imagine that other things are conscious.

So this pragmatic reconsideration of metaphor and the computational implementation of the redesigned model offer at least the beginning of an explanation for the mind's propensity to figuratively project its own consciousness onto the entities that it encounters in the world. This final observation regarding the relationship between metaphor and consciousness can be turned into a possible stance in the debate regarding the controversial construct that reimagines the mind as a computer: if anything, it is the mind that projects consciousness onto the computer, not the computer that stands in as a model for what the mind does. The conceit of the mind as computer seems to easily forget that the operations of a computer are only meaningful by virtue of the values assigned to its inputs and outputs by some agent who is plugged into reality in a deeply intentional way—but then the mysteriousness of consciousness likewise evades the question of what exactly it is that is doing the conscious sensing, leaving only the fanciful notion that all nature of other things can consciously sense, as well. And so in the end, the metaphor of the mind as a computer is perhaps actually just a reversal of the metaphor of a computer as a kind of mind, a lending out of the self which is actually just a specific case of what conscious minds, in their incessant and incurable projecting, do to everything in the world.

ACKNOWLEDGEMENTS

REFERENCES

[1] Aristotle, *The Poetics*, Macmillan and Co, London, 1895.
[2] Merrie Bergmann, 'Metaphorical assertions', *The Philosophical Review*, **91**, 229–245, (1982).
[3] Max Black, 'Metaphor', in *Proceedings of the Aristotelian Society*, volume 55, pp. 273–294, (1955).
[4] Max Black, 'More about metaphor', in *Metaphor and Thought*, ed., Andrew Ortony, 19–41, Cambridge University Press, 2nd edn., (1977).
[5] William Blacoe and Mirella Lapata, 'A comparison of vector-based representations for semantic composition', in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 546–556, (2012).
[6] Peter Carruthers, *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford University Press, 2011.
[7] Robyn Carston, 'Metaphor: Ad hoc concepts, literal meaning and mental images', in *Proceedings of the Aristotelian Society*, volume CX, pp. 297–323, (2010).
[8] Robyn Carston, 'Metaphor and the literal/nonliteral distinction', in *The Cambridge Handbook of Pragmatics*, eds., K. Allan and K. M. Jaszczolt, Cambridge University Press, (2012).
[9] David J. Chalmers, *The Conscious Mind*, Oxford University Press, 1996.
[10] Andy Clark, 'Language, embodiment, and the cognitive niche', *Trends in Cognitive Sciences*, **10**(8), (2006).
[11] Donald Davidson, 'On the very idea of a conceptual scheme', in *Proceedings and Addresses of the American Philosophical Association*, volume 47, pp. 5–20, (1974).
[12] Donald Davidson, *Inquiries into Truth and Interpretation*, chapter What Metaphors Mean, Clarendon Press, Oxford, 2nd edn., 1978.
[13] Donald Davidson, 'A coherence theory of truth and knowledge', in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, ed., Ernest LePore, (1986).
[14] Daniel C. Dennett, *Consciousness Explained*, The Penguin Press, London, 1991.
[15] Daniel C. Dennett, *Intuition Pumps and Other Tools for Thinking*, W. W. Norton and Company, 2013.
[16] Fred I. Dretske, *Knowledge and the Flow of Information*, CSLI Publications, 1981.
[17] Daniel L. Everett, 'Cultural constraints on grammar and cognition in pirahã: Another look at the design features of human language', *Current Anthropology*, **46**(4), 621–646, (2005).
[18] Jerome Feldman, 'Embodied language, best-fit analysis, and formal compositionality', *Physics of Life Reviews*, **7**(4), 385–410, (2010).
[19] Luciano Floridi, *The Philosophy of Information*, Oxford University Press, 2011.
[20] Peter Gärdenfors, *Conceptual Space: The Geometry of Thought*, The MIT Press, Cambridge, MA, 2000.
[21] Andrew Gargett and John Barnden, 'Gen-meta: Generating metaphors using a combination of ai reasoning and corpus-based modeling of formulaic expressions', in *Proceedings of TAAI 2013*, (2013).
[22] Raymond W. Gibbs, Jr., *The Poetics of Mind*, Cambridge University Press, 1994.
[23] James J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Miffline, Boston, 1979.
[24] Mary Hesse, 'The cognitive claims of metaphor', *The Journal of Speculative Philosophy*, **2**(1), 1–16, (1988).
[25] Mary B. Hesse, *Models and Analogies in Science*, Sheed and Ward, New York, 1963.
[26] Julian Jaynes, *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, Penguin Books, 1976.
[27] Paul Kay and Luisa Maffi, 'Color appearances and the emergence and evolution of basic color lexicons', *American Anthropologist*, **101**(4), 743–760, (1999).
[28] George Lakoff and Mark Johnson, *Metaphors We Live By*, University of Chicago Press, 1980.
[29] Stephen C. Levinson, 'Yélî dnye and the theory of basic color terms', *Journal of Linguistic Anthropology*, **10**(1), 3–55, (2001).
[30] Dmitrijs Milajevs, Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, and Matthew Purver, 'Evaluating neural word representations in tensor-based compositional settings', in *Proceedings of EMNLP 2014*, (2014).
[31] Andrew Ortony, 'Why metaphors are necessary and not just nice', *Educational Theory*, **25**(1), 45–53, (1975).

[32] *Metaphor and Thought*, ed., Andrew Ortony, Cambridge University Press, 2nd edn., 1993.

[33] Charles Sanders Peirce, *Peirce on Signs*, The University of North Carolina Press, Chapel Hill, NC, 1991.

[34] Richard Rorty, 'Unfamiliar noises: Hesse and davidson on metaphor', *Proceedings of the Aristotelian Society*, **61**, 283–296, (1987).

[35] John R. Searle, 'Metaphor', in *Metaphor and Thought*, ed., Andrew Ortony, Cambridge University Press, (1979).

[36] Alfred Tarski, *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, Hackett, Indianapolis, IN, 2nd edn., 1983. Translated by J. H. Woodger.

[37] Peter D. Turney and Patrick Patel, 'From frequency to meaning: Vector space models of semantics', *Journal of Artificial Intelligence Research*, **37**, 141–188, (2010).

[38] Josef van Genabith, 'Metaphors, logic and type theory', *Metaphor and Symbol*, (2001).

[39] Tony Veale, 'Dynamic type creation in metaphor interpretation and analogical reasoning: A case-study with wordnet', in *Conceptual Structures for Knowledge Creation and Communication*, 146–159, Springer, (2003).

[40] Dominic Widdows, *Geometry and Meaning*, CSLI Publications, Stanford, CA, 2004.

A Formal Model of Metaphor in Frame Semantics

Vasil Penchev[1]

Abstract. A formal model of metaphor is introduced. It models metaphor, first, as an interaction of "frames" according to the frame semantics, and then, as a wave function in Hilbert space. The practical way for a probability distribution and a corresponding wave function to be assigned to a given metaphor in a given language is considered. A series of formal definitions is deduced from this for: "representation", "reality", "language", "ontology", etc. All are based on Hilbert space. A few statements about a quantum computer are implied: The so-defined reality is inherent and internal to it. It can report a result only "metaphorically". It will demolish transmitting the result "literally", i.e. absolutely exactly. A new and different formal definition of metaphor is introduced as a few entangled wave functions corresponding to different "signs" in different language formally defined as above. The change of frames as the change from the one to the other formal definition of metaphor is interpreted as a formal definition of thought. Four areas of cognition are unified as different but isomorphic interpretations of the mathematical model based on Hilbert space. These are: quantum mechanics, frame semantics, formal semantics by means of quantum computer, and the theory of metaphor in linguistics.

1 INTRODUCTION

The thesis of the paper is fourfold: (1) Metaphor can be seen as the interaction of at least two frames in a sense of frame semantics. (2) Then representation can be interpreted as the particular case of zero interaction between the frames. (3) In turn, this allows of the frames to be interpreted formally as correspondingly "reality" and the "image of reality", and language as an (even one-to-one) mapping between those two universal and formal frames of "reality" and its "image". (4) Metaphor can be further represented formally as the "entanglement"[2] of two or more frames and thus in terms of quantum information.

That thesis has advantage (or disadvantage from another viewpoint) to be self-referential and paradoxical: Indeed the so-defined concept of metaphor is in turn the interaction between two frames: both that of frame semantics and that of formal semantics and consequently it would be "only" a metaphor if the frame semantics and formal semantics can interact as this text advocates; and vice versa: if any scientific notion is expected to be a representation of reality, this text should be zero-content for the set of its extension should be empty. Nevertheless, that explicit paradox is rather an advantage as the analogical paradox generates the development of language and thus perhaps this text as a live part of it.

The mathematical formalism of quantum mechanics (the so-called quantum mathematics) can serve for a formal theory of metaphor and thus for a serious technical formulation applicable to AI. However, the demonstration of the latter is absolutely impossible in the volume of the present paper. Its purpose is restricted only to *outlining the possibility* of a "quantum theory of metaphor".

That "quantum theory of metaphor" can be defined as that mathematical model of metaphor, which is based on Hilbert space very well utilized already by quantum mechanics.

Thus the suggested "quantum theory of metaphor" would share a common mathematical formalism with quantum mechanics. If that is the case, the representation of metaphor in terms of quantum mechanics is neither merely a loose analogy nor any metaphor of "metaphor", but rather a mapping between two different interpretations of the underlying model of Hilbert space.

Furthermore, the notions, approaches and even visualisations of quantum mechanics are exceptionally well developed in detail. They allow of that theory of metaphor called quantum to be represented immediately by a complete language including both mathematical model and huge practical area such as quantum mechanics.

Some of the most essential concepts of quantum mechanics necessary also to that theory of metaphor are "entanglement", "quantum information", and "quantum computer" defined bellow. Besides them, still a few terms need some specification, namely: "frame semantics", "frame" "formal semantics":

"Frame semantics" is meant in the sense of Charles J. Fillmore: "Frame semantics offers a particular way of looking at word meanings, as well as a way of characterizing principles for creating new words and phrases, for adding new meanings to words, and for assembling the meanings of elements in a text into the total meaning of the text" [1].

"Frame": "The idea is that people have in memory an inventory of schemata for structuring, classifying and interpreting experiences, and that they have various ways of accessing these schemata and various procedures for performing operations on them" [2]. "By the term 'frame' I have in mind any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits ..." [1]. The "frame" already linked to formal semantics is specified as a set of well-orderings referring to something as its "logic", in which any property, relation, part or feature of that something can be understood by somebody or by some group. Consequently, that formal and semantic "frame" means the relation between the wholeness of that something and the "logic" of it as a collection of well-orderings.

---

[1] Dept. of Logical Systems and Models, Institute for the Study of Societies and Knowledge at the Bulgarian Academy of Sciences, Sofia, Bulgaria, email: vasildinev@gmail.com.
[2] Entanglement can be interpreted as a kind of interaction due to wholeness: If two or more entities constitute a common system, they can interact with each other by the whole of the system itself, i.e. holistically, rather than only by some deterministic and unambiguous mechanism.

"Formal semantics" is a term used both in logic and in linguistics[3] but in partially different meanings [3]. The common is the utilization of mathematical and logical models. However, the logical "formal semantics" addresses the natural entailment in language in terms of logical sequence while the linguistic "formal semantics" discusses rather the correspondence both of linguistic units and the wholeness of texts to reality in terms of mathematical mappings, set theory, and logic [4, 5]. These meanings will be "entangled" in this paper by the mathematical concept of well-ordering, which can refer both to any logical sequence, and thus to any entailment in language, and to set theory including the axiom of choice, and thus to any one-to-one mapping of language and reality, such as a presentation.

"Entanglement" is a term in quantum mechanics, meaning the information interaction between two or more quantum systems and thus being fundamental for the theory of quantum information. However, the formal and mathematical definition of "entanglement" as that Hilbert space[4], which cannot be factorized to any tensor product of the Hilbert spaces of subsystems, allows of the term to be generalized to any model utilizing Hilbert spaces. For the formal and semantic model used here is based on Hilbert space(s), the concept of entanglement is applicable. It is the mathematical base for the model of metaphor.

"Quantum information" is a term initially coined by quantum mechanics to describe the base of a generalized kind of information underlying all quantum mechanics. So, quantum information can be interpreted as both transfinite series of bits and finite or infinite series of qubits. A bit is the elementary choice between two equally probable alternatives, and a qubit (i.e. quantum bit) can be interpreted as the elementary choice among an infinite set of alternatives though it is initially defined in quantum mechanics as the normed superposition of two orthogonal subspaces of Hilbert space. The quantity of information whether classical or quantum is the quantity of the corresponding elementary choices (whether bits or qubits) necessary for transforming a well-ordering to another (both, whether finite or transfinite). Thus quantum information can be interpreted as the quantity of elementary choices necessary to transform a frame into another and consequently the information of any metaphor formalized as above.

"Quantum computer" [7, 8, 9] is a mathematical model involved by quantum mechanics to interpret its formalism as a generalized kind of calculation, processing quantum information. Thus all physical states and processes may be also seen as computational.

The advantages of the suggested theory of metaphor would be the following:

It relies on a developed and utilized model though in a rather different scientific area.

It can be applied practically as this is sketched (only roughly) in Section 2.

It would aid the formal reconstruction of semantic interactions as a whole as well their historical change by investigating the correlations in the uses in texts and discourses.

It allows of far reaching unifications, generalizations, and philosophical conclusions.

A section (6) is devoted to the unity of thesis as a single, coherent and contextual whole consisting of the distinguished parts (namely the four "folds" of the fourfold thesis above). The mathematical model lent by quantum mechanics is the common base.

Nevertheless some ideas can be considered in their own right even out of the model, e.g. representation as a particular, borderline and limiting case of metaphor.

However this seems to be impossible as to others, e.g. the converse relation of model and reality, proposed near the end of Section 4. Those are logical corollaries from the utilized model.

The argumentation for the thesis has four corresponding points:

(1) Metaphor can be understood as the appearance of a new frame by interaction of two or more initial frames for some essential part of each of them is shared by all. Thus the understanding of each of them separately generates immediately the understanding of the metaphor as a new whole [10, 11] demonstrating therefore the appearance of a new frame, which is not the simple additivity of the sub-frames composing it. The set of well-orderings formalizing semantically a frame can be substituted by a point of Hilbert space [12], and interpreted as a wave function[5] of a quantum system [13]. Any possible frame is measurable as a single value of quantum information. Then the metaphor will be interpretable as the entanglement of the quantum systems corresponding to each sub-frame composing it.

(2) Representation can be interpreted after that as a particular and borderline case of metaphor, a "zero" metaphor, or just as the simple additivity of the sub-frames composing it. The corresponding wave functions are orthogonal to each other and there is no entanglement between them.

(3) Language is reduced to an infinite countable set (A) of its units of meaning, either words or propositions, or whatever others [14]. It includes all possible meanings, which can be ever expressed in the language rather than the existing till now, which would always a finite set. The external twin of reality is introduced by another set (B) such that its intersection with the above set of language to be empty. The union of them (C=A∪B) exists always so that a one-to-one mapping (f: C↔A) should exist under the condition of the axiom of choice. The mapping (f) produces an image (B (f)) of the latter set (B) within the former set (A). That image (B (f)) serves as the other twin of reality to model the reality within the language as the exact representation [15] of the reality out of language (modelled as the set B). In the model, the necessity and sufficient condition of that representation between reality both within and out of the language is just the axiom of choice: If the axiom of choice does not hold, the relation between the sets B (f) and B cannot be defined rigorously as an exact representation but rather as some

---

[3] Some authors doubt the relevance of formal semantics to natural languages [6].

[4] The complex Hilbert space is the fundamental mathematical structure underling quantum mechanics. It is a vector space defined over the field of complex numbers. Hilbert space can be thought as the infinitely dimensional generalization of the usual three-dimensional Euclidean space where furthermore the real numbers are replaced by complex ones. Just the complex Hilbert space is meant for "Hilbert space" in the paper. It allows of: arithmetic and geometry to be generalized and thus unified into a single structure; the possible and actual to be not more than different interpretations of a single mathematical structure.

[5] The term "wave function" is used bellow without quotation marks also a synonym of an element of the complex Hilbert space. Exactly speaking, the former is the common interpretation of that element in quantum mechanics.

simile and the vehicle between the two twins of reality can be only metaphor[6].

(4) Metaphor formalized as above is representable as the wave function of the frame compounded by two or more sub-frames, which interact between each other by means of the shared nonzero intersection. The quantity of quantum information of a metaphor is different from that quantity of the corresponding representation. Thus the metaphor demonstrates the entanglement of the composing sub-frames after they have been formalized as points in Hilbert space [16].

The intuitive sense for metaphor to be represented as the entanglement of its terms is the following. The meaning of any term in a metaphor influences the meanings of the rest.

Consequently, their meaning within the metaphor is essentially different from those of the terms by themselves.

Any mathematical model of metaphor needs a certain relevant quantity of that influence. Once that model involves Hilbert space(s), the entanglement and the corresponding quantity of quantum information are the most natural applicant for describing the degree of that influence.

However, the metaphor itself being already mathematically modelled serves to describe the degree of entanglement between different formal realities (or "languages") in Section 4 and Section 5. Then the formal concept of language is accordingly generalized from a simple representation of reality, i.e. its identical "twin", to a metaphorical image of both reality as a whole and its separate elements such as "things".

The paper is organized as follows. The sections from 2 to 5 argue for the four "folds" of the thesis: (1) to (4) above. Section 6 unites them into a single viewpoint. Section 7 presents the conclusions and provides directions for future work.

2 METAPHOR AS INTERACTION OF FRAMES

Metaphor can be seen as the interaction of two or more frames as follows. Any frame corresponds of some unit of meaning such as a word. The meaning is understood as a whole, i.e. all links between this unit and other units in the frame are actually given according to frame semantics. One can suppose language as the maximal frame containing all other frames as sub-frames. Anyway the most part of language remains absolutely or almost irrelevant to the understanding of any given term. The other, quite small part most relevant to the understanding can be used for its definition. Consequently, the understanding of a meaning can be thought as an exactly determined position in the maximal frame of language, in which the neighbour links are crucial, the next links are less crucial, and the significance of further links weaken very fast, but gradually, moving away from the position

in question and converging to zero as to the most part of the language [17].

The same picture can be repeated for arbitrarily many meanings, and particularly for one more:

Let us figure that both meanings are simultaneously active and their joint understanding is supposed. If both meanings are neighbour or at least relevant in definition, this is rather a proposition than a metaphor. The link between them is explicit in the frame of each of them.

However that is not the case of a proper metaphor where the link connects two areas, each of which is relevant for the understanding of one term, but irrelevant for the other one.

Obviously, the transition between the compound frame of a proposition and that of a metaphor is gradual [10].

Metaphor can be seen as a generalization of proposition referring to remote meanings in the maximal frame of language. Proposition does not generate any radically new meaning irredundant to those of its parts. The meaning of a proposition can be called "analytical" in a *broad and linguistic sense*[7].

Any metaphor appeals to some implicit meaning relevant to the pathway frame between the connected ones. However, that pathway frame of a metaphor is not objective. It depends not only on the connected frame, but also on the person(s) who understand(s). The pathway and thus the implicit frame are not unambiguously determined: it includes also the personality and biography of who understands. The meaning of a metaphor can be called "synthetic" in a broad and linguistic sense:

One can utilize the picture of the maximal frame, in which are chosen two positions as two points. Furthermore, the proposition connects them by a single "classical trajectory" while, the metaphor does the same by *all possible trajectories*, each of which is differently probable. Any understanding chooses only one of them. The mapping analogy to the Feynman interpretation[8] of quantum mechanics [18, 19, 20, and 21] is obvious. It addresses further the idea for the mathematical formalism of quantum mechanics to be only adapted to the relevant terms of frame semantics:

Indeed any measurement in quantum mechanics corresponds to a given understanding of what the metaphor mean. The metaphor unlike any proposition does not predetermine how it should be understood, however it defines implicitly a wave function of all possible understandings as the set of pathways, in any of which it can be interpreted equally justifiably.

Entanglement and the Feynman interpretation are both deduced from the mathematical formalism, but historically independent of each other. Nevertheless, there exists the following rigorous logical link between them:

The Feynman interpretation implies entanglement:

---

[6] The axiom of choice is independent of the other axioms of set theory in the usual systems of its axioms. The former case corresponds to the systems with the axiom of choice, the latter without it. However in fact, the utilized model of Hilbert space is invariant to it without being independent of it in a sense: Quantum mechanics uses Hilbert space both with and without the axiom of choice in two interpretations of quantum mechanics, which identify to each other and anyway distinguish from each other. This is rather a special and inherent property of Hilbert space than an accidental one brought in by quantum mechanics for interpretation.

[7] That "broad and linguistic sense" means that the proposition is a series, the elements of which are ordered in a whole. Anyway this is not the rigorous formal and logical deduction, which is analytical in a *narrow sense* for the premise implies the conclusion necessarily. The analyticity of a proposition is pragmatic and due to the possibility and probability of a rather expected link being usual and more or less often used. Metaphor is rather unexpected and nevertheless understandable.

[8] The essence is any motion or change to be generalized as done in infinitely many paths simultaneously rather than in a single one. The metaphor can be thought in the same way as the motion from a term to another or others in "many paths", each of which is an interpretation of the metaphor in questions and can be realized by somebody.

Indeed any "path" between two or more quantum entities means that they share at least one of their own possible states as common. And vice versa: if there is not entanglement, the Feynman interpretation would be impossible for this means that the entities are orthogonal to each other and thus they are not able to share any common states.

Furthermore, the exact mathematical formalism, which the Feynman interpretation implies, considers Hilbert space only as an approximation or as a limit after *infinitely* many "paths". In fact, that approximation and thus the nonzero difference between Hilbert space and the proper formalism of that kind are inherently necessary for that interpretation because this allows of entanglement to "sneak" implicitly into it.

Consequently, the Feynman interpretation is a stronger statement than the standard mathematical formulation about single, independent and thus non-entangled Hilbert spaces, which are all equivalent to a single Hilbert space[9].

Once the Feynman interpretation is involved for the mathematical model of metaphor as above, this implies immediately that entanglement is also though implicitly introduced and should be discussed in the framework of that model.

The Feynman interpretation further means that if it is universal, all quantum systems are entangled, and the standard consideration of quantum mechanics by single and non-entangled Hilbert space is not more than a working idealization and simplification.

That states of affairs in quantum mechanics can be forthwith interpreted in terms of the utilized model of metaphor: Representation is not more than a working idealization and simplification of metaphor: one statement, which will be discussed in detail in the next section.

The situation of two terms can be continued to more than two, even to arbitrarily many, and one is able even to consider the case of the metaphor of metaphors [22] as well that of the "proposition of metaphors". The method for that continuation is the relevant interpretation in terms of quantum mechanics in order to be borrowed the very well developed mathematical model.

Practically, one needs some relevant, reliable, and relatively unambiguous method for any given metaphor in a given language with its use and history to be adequately determined its wave function. This method can involve the following stages:

1. Determining a broad set of associative series, which can connect the terms of the investigated metaphor.

2. Structuring this set as a directed graph [23].

3. Determining the combinatory frequency of each vertex in the entire dictionary of the language or in any as contemporary as historical sub-dictionary if need be.

---

[9] However one has to mean that any quantum system referring to a single Hilbert space can be always exactly and equivalently represented as consisting of two or more entangled subsystems and correspondingly Hilbert spaces. Then the viewpoint of the system differs from that of any subsystem. The Feynman interpretation is a way the viewpoint of the quantum whole to be represented as a certain function (namely its wave function) of the viewpoints of its virtual classical "parts", each of which is featured by a single classical "path". The suggested model of metaphor being considered as a whole would consist of the virtual parts of its interpretations, any of which is featured by its own proper associative path and a corresponding probability of this path calculable by relevant frequency uses.

4. Calculating the frequency and probability in any possible pathway in the graph.

5. Summarizing these data as a probability distribution.

6. Approximating this probability distribution [24] by a wave function.

7. Eventually interpreting and modelling this wave function as a state of a quantum system and thus of a quantum computer.

Only stage 1 depends crucially on the human creativity to be figured all thinkable and unthinkable associative series, which can connect the terms of a metaphor. All rest stages can be accommodated for relevant software.

However, ever this first stage might be replaced by a formal frequency use analysis of common terms in the frames of all terms constituting a given metaphor. One should consider those frames as frequency use in the context of a given term and consisting of two, three, four and so on words. Consequently, the following stages «1'» and «1″» can substitute the above «1»:

1': Formally determining the frame of each term constituting the given metaphor as frequency uses of two, three, four, five, and so on words, containing the term in question.

1″: Determining the frequency use of common terms in the frames of the terms of the investigated metaphor.

Those stages can be quite roughly illustrated by an imaginary example for their application about a real metaphor, e.g. "The moon is sad".

First of all, this is an obvious metaphor, which connects a celestial body, which is impossible to be sad, with a human mood, that to be sad: Who is sad cannot be anything inanimate such as the moon.

Furthermore, "Google" shows that the exact phrase as above is used in 59,000 web sources (retrieved on 14.03.2015). Nevertheless, the phrase is found in no case in the huge data base of English literature in "Ngram Viewer" of "Google books" (again then). Consequently, this is a real contemporary metaphor rather than a "white metaphor" coining Derrida's metaphor about any too used metaphor.

There are at least two different practical methods, which would give also different results perhaps, to be determined the paths and their corresponding probabilities for the latter term, "sad", to be reached starting from the former term, "moon".

The one method would construct the frames of both terms by means of main frequency uses of small contexts containing the terms and would search for coincidences of terms belonging to both frames.

One can figure as an imaginary example that the pair (moon, round) has frequency use "f_1" and probability "p_1" calculable as the ratio of "f_1" to the number of all considered frequency uses in the frame of "moon". Furthermore, the triple (round, face, sad) is analogically featured by "f_2" and "p_2" in the frame of "sad". "Round" is the searched coincidence. It allows of constructing some relevant function "$P_1 (f_1, p_1, f_2, p_2)$", which would suggest a value of the composed path (moon, round, face, sad) connecting both terms of the metaphor in a possible way.

The other method would consider only the frequency uses of those pairs, the series of which starts from "moon" and finish to "sad".

In the above example, those would be: (moon, round), f_3, p_3; (round, face), f_4, p_4; (face, sad), f_5, p_5. They would imply some $P_2 (f_3, p_3, f_4, p_4, f_5, p_5)$ of the same path however calculated by the latter method.

If that procedure either in the former or in the latter method is repeated as to many enough paths, one can yield the probability distribution, which refers to the metaphor "The moon is sad" in English, with any preliminarily defined exactness. Then, the characteristic function of that probability distribution will represent the searched wave function of the metaphor in question.

The above two methods can be further modified and mixed in different proportions. However, they reflect two different ways for the model of metaphor to be understood: either as the entanglement of the frames of terms constituting a given metaphor or as a single frame of the metaphor as a whole, which is practically reduced to a set of series corresponding to paths between the terms of the metaphor.

Anyway the goal of the paper is only the possibility in principle as well as a schematic diagram of how the metaphors first interpreted in terms of frame semantics to be further modelled mathematically and then computationally.

3 REPRESENTATION AS A PARTICULAR CASE OF METAPHOR

The next step refers to representation: How the representation to be grounded on metaphor? The usual way is the reversed: How the metaphor to be founded by representation, which is granted as a self-obvious base?

However, the above mapping to quantum mechanics leads just to the metaphor to be the starting point. The end point is not so the representation by itself, but the concept of reality to be obtained in a formal and mathematical way [25] in order to be modeled.

The representation can be considered as a particular and borderline case of metaphor following the method for quantum mechanics to be reduced to classical mechanics by the principle of correspondence.

The problem is the following. Some metaphor is given. Which are the boundary conditions, on which its wave function can be transformed into that of a corresponding representation? The wave function of a representation is degenerated in a way so that the corresponding probability distribution is reduced to a single infinite pick in a single point, i.e. to the Dirac δ-function.

That result for the probability distribution in all associative ways of the metaphor in question can be obtained so: the interval of nonzero probabilities converges to the limit of a single point.

The process of convergence requires both decreasing the associative "distance" between the connected terms of the metaphor (which are at least two) and increasing the extension of the generalization of the terms so that the set of all associative pathways to be able to be reduced gradually to a single one. If that is the convergence, the corresponding directed graph of the metaphor will degenerate to a directed segment and even to a directed segment of zero length. The latter in turn is equivalent to a bit of information [26]: the "cell" of the segment possesses two equally probable, but alternative state of each of the two ends.

This would correspond to the degenerated or "ontological" metaphor: '"A" is A' decodable as the dialectic judgment that both "A" is A, and "A" is not A. The two ends of the "zero segment" are: "A" and A (whatever A is).

The directed segment of zero length (or a bit) means an elementary choice as well as an identical mapping. If these concepts are applied to an infinite set, they require the axiom of choice and even a special case of invariance in relation to it. That invariance consists in this, any subset of any set not only to be able to be enumerated by virtue of the axiom of choice, but also the set and the enumerated image of it to be identified.

The mathematical model of representation deduced from the metaphor should include all aforesaid formal properties.

Let us now interpret these mathematical features of representation in terms of frame semantics, i.e. as an interaction between two frames, which relation can be even identical. That interaction is zero in both opposite cases: both where the frames are absolutely independent of each other and where they coincide.

Even more, both cases can be identified by the above formal properties of representation as the "two ends of a directed segment of zero length" or as the "ontological metaphor": '"A" is A'.

Then the "class of all representations" can be defined as 'reality' in terms of the formal frame semantics. Reality can be deduced from representation, which in turn can be deduced from metaphor.

The formal and mathematical concept of reality is crucial for modeling any intellect able to be standalone. The demarcation line between a machine however "clever" and an intellect however "stupid" is just the concept of reality, which is inherent for the latter and somebody else's for the former. Thus the machine however "intelligent" remains a machine in somebody else's reality, e.g. a human being's.

Reality equivalent to the class of all representations is equivalent also to the aforesaid invariance to the axiom of choice for the class of all representations coincides with that invariance. However, it can be defined only on infinite sets.

Practically, this means that the formal concept of reality defined as above can be modeled only by some quantum system, i.e. on a quantum computer rather than on a Turing machine (i.e. on any standard computer independent of its power) always representing always a finite series after finishing effectively by any result .

A representation modeled on a quantum computer is a measurement of it. Any direct measurement means for a quantum computer to be irreversibly demolished, though:

This means that the superposition of all possible states, which is essential for its definition, is reduced to a single one, namely what is measured. Indeed the processing of a quantum computer consists in a reversible and smooth change of all elements of a set of probability distributions. Thus the statistical probabilities of the corresponding ensemble of measured results are changed as the output of that computer. However, the measurement of any state cancels irreversibly its work and it is destroyed in fact.

Consequently, the attempt to be modeled that formal concept of reality on a quantum computer fails for the set of representations, i.e. measurements are not infinite: even if the measurements are done of a collection of quantum computers. Furthermore, that collection is not only finite, but also a statistical ensemble rather than a coherent state.

One has to search for other, nondestructive ways for mappings of a coherent state into another or other of a quantum computer rather than into the elements of a statistical ensemble.

This requires the correspondence of reality and image to be first reformulated in a generalizing way allowing of the communication between them by means of entanglement.

4 HILBERT SPACE: REALITY AND ITS MAPPING WITHIN A QUANTUM COMPUTER

The next step refers to the formal concept of language again by means of Hilbert space [27, 28]. The goal of that step addresses reality to be generalized in way allowing of sharing reality not to lead to demolishing the quantum computer. The constraints and quantitative laws of that sharing are further problems.

Once reality is defined formally as a special set of mappings, one can continue generalizing to broader and broader sets of mappings. They can be also considered as "languages" mapping the so defined "reality" in different ways. Furthermore, each that language offers a different metaphor in general[10] for each "element of reality" being a representation. Then any collection of metaphors about those "elements of reality" is a language obviously defined already formally.

In other words, the language is defined as a particular set of primary (or "elementary") metaphors, in which at least one term is necessarily an "element of reality" while the others designate or define it. Two frames correspond to them in frame semantics being linked to each other by a wave function, i.e. by a point in Hilbert space according to the model introduced in section 2.

This means that any language should be consider as a state of the quantum field over reality. The term of "quantum field" is meant as usual in quantum mechanics, i.e. as a mapping of a set (the set of all representations, or "reality") into Hilbert space.

The "set of all possible states of the so-defined quantum field" including all possible languages will be designate as 'ontology'[11].

Consequently, the concept of ontology is implied much broader than that of reality. If any image of reality in any language is interpreted as another reality, then ontology is the class of all realities or of all possible worlds.

One can demonstrate that those formal concepts are able to be modelled entirely within Hilbert space in a quite natural way. Indeed "representation" corresponds to the relation of two coinciding elements of the two dual spaces. They are both identical and complementary.

Consequently, the so-defined formal concept of reality is inherent to Hilbert space. If Hilbert space is considered as a model shared e.g. by quantum mechanics, that reality is internal rather than external to it. It is complete to that reality.

The interrelation of model and reality (more exactly, the so-defined reality as a formal model) is rather extraordinary in comparison with classical physics, science, and epistemology, being "reversed" in a sense. Model contents the model of reality rather than reality contents the reality of model.

Then any language is a mapping of Hilbert space [29] into itself, and thus any physical quantity[12] is a language defined formally as above (but not vice versa).

Furthermore, Hilbert space can be considered as a quantum computer, and any point in it as a state of it. So that quantum computer should content reality in the sense of the above formal model of reality within itself being therefore standalone rather than a machine within somebody else's reality.

---

[10] Particularly some metaphors in some languages can coincide.

[11] T. Giraud offers a fundamentally different ontological perspective [32].

[12] In the way as it is defined in quantum mechanics.

However, there is a considerable problem of how two or more different realities are able to communicate. Particularly, how is a quantum computer able to transfer a result to us without demolishing itself and thus destroying also that other reality within it and different from ours?

As we will see: only "metaphorically".

5 METAPHOR IN TERMS OF ENTANGLEMENT

The next step requires the relation of any two "languages" to be defined in terms of Hilbert space(s) therefore involving entanglement between them. The goal is: some nondestructive way for transmitting information between two or more realities identified as languages to be outlined. The way of measurement has already excluded above as destructive.

Let there are two different "metaphors" of one and the same "element of reality" in two languages, i.e. two wave functions. The "element of reality" can be excluded and any of the two metaphors can be directly referred to the language (reality) of the other. Those language and reality in the neighborhood of the metaphor are unambiguously defined by the corresponding wave function. Thus the metaphor will "seem" or "appear" as the entanglement of both wave functions from the viewpoint of each of the languages.

One can compare the formal definition of a metaphor in Section 2 as a single wave function with the present definition as the entanglement of two ones. Obviously, these definitions do not coincide: There are two different definitions of one and the same metaphor therefore each one needing some different, but relevant interpretation:

The metaphor defined as in Section 2 as a single wave function should be interpreted as that in the common system of the language or in the universal reality to the particular realities of each term.

The metaphor defined as here, in Section 5 as the entanglement of two or more wave functions should be interpreted as seen from the particular viewpoint of each term of it and thus in the corresponding particular reality.

However, that mismatch is just the nondestructive way for a quantum computer to transmit a result, as we see, only "metaphorically". The transfer is "less metaphorical", i.e. more precious, the quantum computer will be more influenced by the transfer, even demolished after any absolutely exact transmission of its result. The mismatch depends on the quantity of entanglement, in particular, on that of the quantum computer and our reality.

If one of the terms of the metaphor is permanent, e.g. anchored in our reality, the change of the others can be interpreted as the metaphorical "message" thus poetically [30, 31]. The quantum computer turns out to be a "poet".

Practically, the transmitted result will be a change of the rest frames to an anchored frame postulated as that of reality as to our reality. That change of a few frames being also a change of metaphor and an arbitrary[13] operator in Hilbert space can be defined as a single elementary thought [33].

Consequently, a quantum computer cannot report a result in a nondestructive way, but can communicate a thought just as a

---

[13] That is neither self-adjoint, nor linear in general.

human being can. If the thought is clearer, the computer is more "obsessed" by it: i.e. its state and thus future work will be more influenced by its communication.

6 THE UNITY OF THESIS

One can deduce the following from summarizing Sections 2 – 5:

From 2: Metaphor can be represented as an interaction of frames in terms of frame semantics, and then modelled formal as a "wave function", i.e. as an element (point, vector) in Hilbert space.

From 3: Representation can be defined as a particular case of metaphor, namely as the directed segment between two coinciding frames with a corresponding probability distribution degenerated to a Dirac δ-function. The set of all representations is a formal definition of reality.

From 4: That reality turns out to be inherent and internal to Hilbert space and thus to any quantum computer. It can be also considered as identical to a formal concept of language. The class of all languages (or "realities") defines formally the concept of ontology.

From 5: A quantum computer can report a result only "metaphorically" or "poetically". The report is more precise, the quantum computer is more influenced; and even demolished in the borderline case of absolutely exact report. That report is a change of a metaphor to an anchored term and can be considered as a formal definition of thought.

Conclusion from 2 to 5: Any quantum computer being furthermore standalone and supplied by reality can think. Human thinking can be exhaustedly modelled by a quantum computer.

The unity of the thesis includes a few heterogeneous fields of cognition: quantum mechanics as a theory of nature, frame semantics as a theory of human thinking, the theory of metaphor and representation as a theory of language, quantum computer as a theory of artificial intellect. The four can share a common mathematical model based on Hilbert space(s). This allows of a uniform and even mathematical description both of thinking whether human or artificial and of states and process whether physical or linguistic. These four can be considered as not more than different interpretations of a single model and thus isomorphic to each other.

7 CONCLUSIONS & FUTURE WORK

This paper shows how one can use the concept of frame in frame semantics to define metaphor as an interaction of frames. The Feynman "many-paths" interpretation of quantum mechanics allows of the metaphor to be represented by a wave function and thus the mathematical model of Hilbert space to be involved.

One can demonstrate a general approach for any given metaphor in any given language to be assigned a relevant probability distribution and then a wave function. Though the approach is shown by the example of two terms, it can immediately extend to more than two terms following the pattern of quantum mechanics: any separate position in the Feynman model corresponds one-to-one with a term of the metaphor.

The formal model of metaphor implies that of representation as a particular and borderline case of the "ontological" metaphor '"A" is A', and the Dirac δ-function as the corresponding probability distribution. This allows of a formal definition of reality as the set of all representations. That reality is inherent and internal to Hilbert space. Thus any quantum computer turns out to be supplied by its inherent and internal reality. Its reality is what guarantees for it to be standalone rather than a machine in somebody else's reality. However, a quantum computer cannot report us any absolutely exact result without self-demolition.

One can define a formal concept of language within Hilbert space as the mapping of "reality", being internal to the Hilbert space, to the same Hilbert space. That mapping can be considered as a quantum field in the standard definition of quantum mechanics. However, it can be also interpreted as a language mapping any element of reality (signified) into another (signifier) by means of that metaphor (sign), the wave function of which is the value of the quantum field for this element of reality. Furthermore one can define 'ontology' as the "class of all languages" and therefore of all realities or all possible worlds.

This allows of another formal definition of metaphor as a compound "sign" (i.e. two or more entangled wave functions) consisting of two or more signs referring to different signifier in different languages, but of a single common signified.

That formal concept of language is a "quantum field" on "reality", i.e. as a mapping of the set of the formally defined reality in Hilbert space into the same Hilbert space. Any "element of reality" is a "signified" mapped by the "sign" of a metaphor (i.e. a wave function) into another (in general) "element of reality" as a "signifier". Any "language" is also interpreted as another and different "reality" again formally defined. 'Ontology' is further defined as the "class of all languages" and thus that of all realities.

The other, new, and different formal definition of metaphor is given as the relation between different signifiers of a single element of reality as a signified and therefore modeled by two or more entangled wave functions corresponding to the sign of each term in each language.

There will be two distinct definitions of one and the same metaphor: as a single wave function according to Section 2 and as a few entangled wave functions according to Section 5. The quantitative mismatch (being due to the entanglement) between the two definitions can be represented back in terms of frame semantics as a change of a frame to another, after which all rest terms will change their position to one anchored to that reality (language) chosen as a reference frame, e.g. ours.

That "frame change" being also a "metaphor change" can be defined as an 'elementary thought' [34].

Any quantum computer can transmit any result in a nondestructive way only "metaphorically" or "poetically" rather than literally, i.e. as an elementary thought. The thought transmits the result more exact, it is more "obsessive" for the computer: that is its state and thus reality is more influenced by the event of transmission. The borderline case of an absolutely exact report of the result is tantamount to its demolition.

One can also say that quantum computer thinks in this sense of transferring a message between realities (or languages) metaphorically. Furthermore, the essence of thought turns out to metaphorical and thus poetical in the frame of the present paper.

The unity of the thesis demonstrates that a single and common mathematical model based on Hilbert space can be shared by four scientific fields: quantum mechanics describing nature; frame semantics describing human cognition; linguistics

describing metaphor and representation; theory of quantum information describing quantum computer.

That unity implies the following five directions for future work. Four ones for each of the four fields enumerated above and still one, the fifth for their synthesis developing the underlying mathematical model.

REFERENCES

[1] C. J. Fillmore. Frame semantics. In: *Linguistics in the Morning Calm*. Linguistic Society of Korea (Ed.) Hanshin, Soeul, Korea: 111-137 (1982).

[2] C. J. Fillmore. Frame semantics and the nature of language. In: *Origins and Evolution of Language and Speech*. S. R. Harnad, H. D. Steklis, J. B. Lancaster (Eds.) (Annals of the NY Academy of Sciences, Vol. 280) New York Academy of Sciences, New York, USA: 20-32 (1976).

[3] J. Woleński. What is Formal in Formal Semantics? *Dialectica*, 58(3): 427–436 (2004).

[4] J. Woleñski. From Intentionality To Formal Semantics (From Twardowski To Tarski). *Erkenntnis* , 56(1): 9-27 (2002).

[5] R. Rogers. A survey of formal semantics. *Synthese*, 25 (1): 17-56 (1963).

[6] A. Galton. Formal semantics: is it relevant to artificial intelligence? *Artificial Intelligence Review*, 2(3): 151-165 (1988).

[7] R. P. Feynman. Simulating Physics with Computers. *International Journal of Theoretical Physics*, 21(6/7): 467-488 (1982).

[8] R. P. Feynman. Quantum Mechanical Computers. *Foundations of Physics*, 16(6): 507-531(1986).

[9] D. Deutsch. Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 400 (1818): 97-117 (1985).

[10] L. J. Cohen. The semantics of metaphor. In: *Knowledge and Language: Selected Essays of L. Jonathan Cohen*. J. Logue (Ed.) Dordrecht, Springer Netherlands: 27-40 (2002).

[11] F. Guenthner. On the semantics of metaphor. *Poetics*, 4(2-3): 199-220 (1975).

[12] A. Y. Khrennikov. *Ubiquitous Quantum Structure*. Springer, Berlin; Heidelberg (2010).

[13] D. Aerts and M. Czacho. Quantum aspects of semantic analysis and symbolic artificial intelligence. *Journal of Physics A: Mathematical and General*, 37 (12): 123-132.

[14] M. Jubien. Formal Semantics and the Existence of Sets. *Noûs*, 15(2): 165-176 (1981).

[15] J. H. Dreher. On the semantics of formal representation. *Philosophia: Philosophical quarterly of Israel*, 8(1): 71-78 (1978).

[16] D. Aerts and L. Gabora. A theory of concepts and their combinations I: The structure of the sets of contexts and properties. *Kybernetes*, 34(1&2): 151-175 (2005).

[17] C. Gauker. Contexts in Formal Semantics. *Philosophy Compass* 5(7): 568–578 (2010).

[18] R. P. Feynman. Space-Time Approach to Non-Relativistic Quantum Mechanics. *Reviews of Modern Physics*, 20(2): 367-387 (1948).

[19] R. P. Feynman and F. L. Vernon, Jr. The theory of a general quantum system interacting with a linear dissipative system. *Annals of Physics* (N.Y), 24: 118-173 (1963).

[20] R. P. Feynman; A. R. Hibbs. *Quantum mechanics and path integrals*. New York: McGraw-Hill (1965).

[21] H. D. Zeh. Feynman's interpretation of quantum theory. *The European Physical Journal H*, 36(1): 63-74 (2011).

[22] M. Bergmann. Metaphor and formal semantic theory. *Poetics*, 8(1-2): 213-230 (1979).

[23] W. J. Hutchins. Semantics in three formal models of language. *Lingua*, 28(1): 201-236 (1972).

[24] P. Bruza, J. Busemeyer , and L. Gabora. Introduction to the special issue on quantum cognition. *Journal of Mathematical Psychology*, 53: 303-305 (2009).

[25] D. Pearce and V. Rantala. Realism and formal semantics. *Synthese,* 52(1): 39-53 (1982).

[26] L. V. Arshinskii. Substantial and Formal Deductions in Logics with Vector Semantics. *Automation and Remote Control*, 68(1): 139–148 (2007).

[27] A. Khrennivov. Classical and Quantum Mechanics on Information Spaces with Applications to Cognitive, Psychological, Social, and Anomalous Phenomena. *Foundations of Physics*, 29(7): 1065-1098 (1999).

[28] D. Aerts and L. Gabora. A state-context-property model of concepts and their combinations II: A Hilbert space representation. Kybernetes, 34(1): 192-221 (2005).

[29] C. Begz. Mathematical Approaches to Cognitive Linguistics. *International Journal of Applied Linguistics & English Literature*, 2(4): 192-199 (2013).

[30] T. Van Dijk. Formal semantics of metaphorical discourse. *Poetics*, 4(2-3): 173-198 (1975).

[31] E. Bruss. Formal semantics and poetic meaning. *Poetics*, 4(14-15): 339-363 (1975).

[32] T. Giraud. Constructing formal semantics from an ontological perspective. The case of second-order logics. Synthese, 191(10): 2115-2145 (2014).

[33] N. Chino. Hilbert space theory in psychology (1) - Basic concepts and possible applications. *Bulletin of the Faculty of Letters of Aichi Gakuin University*, 28: 45-65 (1998).

[34] A. Ortony. Metaphor, language, and thought. In: *Metaphor and thought*. A. Ortony (Ed.) Cambridge University Press, Cambridge; New York: 1-16 (1979).

How can metaphors be interpreted cross-linguistically?

Yorick Wilks[1]

Abstract. Research on metaphor as a phenomenon amenable to the techniques of computational linguistics received a substantial boost from a recent US government (the iARPA agency) funding initiative that set up a number of teams in major universities to address the issues of metaphor detection and interpretation on a large scale in text. Part of the stated goal of the project was to detect linguistic metaphors (LMs) computationally in texts in four languages and map them all to a single set of conceptual metaphors (CMs). Much of the inspiration for this funding was the classic work (Lakoff and Johnson, 1980) which posited a set of universal metaphors used across cultures and languages.

I wish to examine the assumptions behind this goal and in particular to address the issue of how and in what representation such CMs can be expressed. I shall argue that a naïve approach to this issue is to make very much the same assumptions as the work of Schank and others in the 1970s (including the present author): namely that there can be a universal language of "primitives" for the expression of meaning, which in practice always turns out to be a form of simple English (or in the case of Schank, atoms like PTRANS, very close to English words). In none of those system was the sense ambiguity of the English-like terms every tackled in a systematic way (though see: Guo 1989). Reviving that assumption for the study of metaphor raises additional issues since, even if the *senses* of the terms in those CM representations could be added, by annotation from a standard lexicon for the CM representations, metaphors often considered to deploy new senses of words which will not be found in existing sense inventories like computational lexicons which, if true, might make such annotation impossible (though later in the paper I shall argue against just that novel deployment of sense in metaphor). This paper is not intended just to present a negative conclusion; I also argue that the representation of metaphors in a range of languages can be brought together within some CM scheme, but that simply reviving the *English-as-interlingua* assumptions of forty years ago is not a good way to make progress in this most difficult area of meaning computation.

In what follows I first discuss first the representation of CMs and ask: in what language are they stated? I argue the need for some inclusion in the representation of the senses of their constituent terms within the CM, or at least a default assumption that the major sense (with respect to some lexicon such as WordNet) is the intended one. I then consider the issue of conventional metaphor and its representation in established lexicons (again such as WordNet) and [1]the effect that can have on detection strategies for metaphor, such as selectional preference breaking.

[1] Florida Institute of Human and Machine Cognition, 15 SE Osceola, Ocala FL 34471. Email: ywilks@ihmc.us

I then argue that the mapping of text metaphors to CMs, as well as the empirical, rather than intuitive, construction of CM inventories requires further use of preference restrictions in lexicons by means of a much-discussed process called projection or coercion. I conclude that only the use of (computable) procedures such as these for metaphor detection and mapping can lead to a plausible program for the large-scale analysis of metaphor in text, and that Lakoff's views on metaphor lack these empirical underpinnings.

1 INTRODUCTION

Understanding prose in any natural language rests first on it being in a language one understands, let us say English for the purposes of this paper. But problems in understanding arise even for native speakers of English as well as with translations, human or mechanical, from other languages. One way of capturing the additional understanding needed that goes "beyond knowing the words and the grammar" is expressed by the term "metaphor". This notion conveniently expresses aspects of culture and figurative expression that go beyond literal or ostensive meaning and are crucial to understanding. These phenomena are sometimes opaque even to those who are experts in the language concerned. Metaphor also has the advantage that it has been an area of research in computer language processing for decades, and one that has yielded real results. That research has been driven in part by the writings of George Lakoff at Berkeley [1] who has developed an approach to metaphor that rests on the following assumptions (in my terms, but I think fairly uncontentious):

- There are similar metaphors found in all cultures that are crucial to understanding language.
- These metaphors can be discovered and listed, even if not exhaustively.
- We can proceed with analysis as if these metaphors can be not only paraphrased but expressed in English.

For example, such a universal metaphor might be expressed (in English) as LIFE IS A JOURNEY and we shall refer to items like this as Conceptual Metaphors (CM). There is then an initial analytic question of how to detect metaphors in text, possibly related to or "expressing" that CM such as *The pensioner was nearing the end of his road*. After locating this sentence as a metaphor there is then the task of matching it to such a stored generalized CM form. We shall refer to linguistic strings like the one in italics as Linguistic Metaphors (LM), There may then be the problem, if one believes in the universal nature of CMs, of how to locate expressions of "similar" metaphors in, say, Farsi to that same CM. The capitalised words in the English form of the CM may themselves have many senses and the question

immediately arises as to how an algorithm is to determine which sense is intended by "LIFE" in that CM: that it is not, say, a "a life as in a children's game of hide and seek, a score token".

One problem with metaphor research, at least from a computational or Natural Langauge Processing (NLP) perspective, is that universal theories like the one above (expressed by the three bullets) have proved resistant to computational implementation, which has not been the case with other, quite different, empirical approaches based on bottom-up detection of LMs in text (e.g. [3], [4]), rather than starting from a set of a priori CMs. We shall now turn to questions about *the representational language in which CMs are stated* and how they to be intuitively understood, since their terms (e.g. LIFE) do not disambiguate themselves

2. THE LANGUAGE OF CONCEPTUAL METAPHORS (CMs)

I shall argue that a crucial aspect of the research problem, which many seem to believe is a solution, is that CMs are classically expressed in English words but without any realization of what that entails. When this is pointed out, a frequent response is that this is an accidental fact of no significance and we can just carry on since though they appear to be English words they are not, but rather some form of symbol outside ordinary natural language. I believe this is profoundly inadequate response. It is in fact a recrudescence of the early discussions in AI and NLP in the 1960s and 1970s on the role of interlinguas in machine translation and in cognitive representations generally. There was a fashion at that time for limited languages (expressed by English primitives terms) within systems for the semantic representation of language content (e.g. in the work of Schank [5]; Wilks, [6] and many others). I am not here defendingthat approach, only pointing out that the extended discussion forty years ago (e.g. in [7]) of the adequacy or otherwise of this limited language of (English-like) primitives to carry the general meaning of language expressions has many similarities to what we are discussing now, nearly fifty years later, in regard to CMs.

There was no real resolution to that controversy of long ago: key references are Pulman's [8] attack on the practice from a linguistic perspective, and Lewis [9]from a philosophical one, in the course of which Lewis invented the term "markerese" for the self-description of language in linguistics (e.g. by Fodor and Katz, [10]) by means of word-like *markers* with no illumination or benefit. But the critiques were not heeded and much such representational work continued, simply because researchers in semantics could see no alternative (outside radical connectionism) to continuing to use symbols to represent the meanings of other symbols. Montague [11] was a philosopher who reacted against markerese but his representations of mean, although more replete with logical forms that those of Fodor and Katz, still were expressed in symbols including English-like words, though now usually expressed in lower case and with an apostrophe attached. Language content had to represented somehow, theorists reasoned, so why not in this English-like language? Dictionaries, after all, describe word meanings using the very language they describe, and so the practice has

continued, ignoring the waves of philosophical and linguistic criticism, simply because there seemed to be no alternative. What has happened is that the language terms used for representation have been embedded in more logical and formal-seeming structures so as to make them palatable, but the underlying issue has not gone away. That issue is: How can I describe semantic content with a term such as MAN, HUMAN or ANIMATE and be confident I know what it means, and not just "means in English"? I shall now turn to how problems of CM representation problems can be ameliorated with the aid of a sense-lexicon.

3. REPRESENTING CMs UNAMBIGUOUSLY WITH MAJOR WORD SENSES

If we are to use CMs at all, no matter how derived or expressed, they must be in as word-sense-neutral a form as we can manage. To my knowledge this has never yet been fully considered as problem, perhaps an insurmountable problem, let alone a solved problem. We cannot just ignore this as we do when we say, for example, that [POVERTY IS A GAP] is a CM, and underlies the metaphor "poverty gap", and that we just know what the senses of the words in the CM are present in that expression and that they make up a CM. Just suppose that we had two CMs in our inventory of universal metaphors that could be written as:

POVERTY IS A GAP

POVERTY IS AN ABYSS

Now suppose we want to locate Russian metaphors and find the text string (LM) containing the keywords : *бедность провал*, which mean roughly "poverty" and "failure". But, and here is the problem "*провал*" can also means "abyss" and "gap" in English; in which case how do we know which of these two so-called universal CMs to match the Russian LM to? Or should we seek for or construct a third CM [POVERTY IS FAILURE]? It seems clear to me that either:

1) The CMs are in some language other than English, in which case how do we know what English word senses the terms above correspond to, since the English words "poverty", "failure" and "abyss" may all have multiple senses in, say, WordNet [12]. If, however, the terms are not English but some universal language of indeterminate syntax and semantics, how can LMs ever be matched to CMs as any serious theory of metaphor seems to require?

2) If however, the terms in the two CMs above *are* in English, and they certainly appear to be, then we need to know what senses those words have in those particular forms, so as to match any word in an English or Russian LM to them.

A natural way of carrying out the requirement in (2) is to tag the English words in the CMs (and the words in any putative LMs) with WordNet senses. Since the EuroWordNet project [12] in which the present author participated, we now have a convenient

way of setting up such a match since that project took the core Princeton WordNet for English as, essentially, an interlingua, and linked senses in the Wordnets for other languages to those core senses. So, for example (and the correctness of these correspondences does not matter for the argument): there may well be an English WordNet sense of "failure", namely failure#1 that is deemed by a EuroWordNet mapping to be the same sense as Провал#1 in the Russian WordNet. Again, there may be a "Провал#3" that similarly corresponds to "abyss#1".

What do we want to say about universal CMs and their ability to support the analysis of metaphor instances in such a case? The first natural thing to say---given the above WordNet assumptions---- is that the original Russian string "*бедность провал*" can express both CMs and we cannot decide which. But that is only true if we cannot decide which sense the last word bears in the Russian LM. If it bears only one of the two noted senses then the Russian LM matches one and only one of the CMs—assuming now the CM terms are tagged with WordNet senses. Russianists should note here that I am ignoring the case issues for the proper expression of that string in Russian and just concentrating on the main forms of the words. Also, I am not suggesting it would be problematic if a LM were to match to two possible CMs, though I do not believe that need be the case here. It could be that other, perhaps pragmatic, factors outside the text would settle the choice. My only point here is that a systematic empirical account of mapping LMs to CMs should take account of this possibility and standard contemporary metaphor theories do not consider the issue at all.

Now a Russian speaker may take that (LM) phrase to have one and only one of those senses in context—assuming the Russian speaker can understand the distinction we are making with the words "failure" and "abyss" in English—let us assume they can, even though the string may be too short and vague for a wordsense disambiguation program to determine the sense in that LM context.

Or, and this is a quite different possibility, is it the case that, in a metaphorical string such as the LM "Poverty is failure" we cannot rely on the normal psychological or computational methods to resolve a word sense for us. Since the content is, more or less, novel, at least on first encounter, the standard disambiguation techniques may well not work because they are all, to some extent, based on redundancy, which does not apply to novel utterances? So, to use an old and hackneyed example, if someone says *The shepherd swung his crook*, we infer that "crook" is a tool for shepherds not a gangster, simply because of the redundant presence of "shepherd". But in LMs this may not be available, unless the metaphor is dead, or lexicalized or otherwise familiar (in which case wordsense disambiguation hardly applies). What I am suggesting is that perhaps in metaphors, especially novel ones, the words must be taken in their basic senses by default, as it were, *because in a metaphor we lack the familiar context to resolve a participating word to any non-basic sense.*

This conclusion is perhaps not very striking but rather obvious: words of a real language, like English, can only function in an interlingua (such as CMs constitute) on condition that they bear their "basic" senses, which will, in WordNet terms, usually mean

#1 for any given word. This implies that in the capitalized English CMs above, each term implicitly has whatever its #1 sense is in WordNet.

So to return to the purported sense correspondence in Euroword-net style:

failure#1 is deemed by a EuroWordNet mapping to be the same sense as Провал#1. Again, there may in addition be a "Провал#3" that similarly corresponds to "abyss#1".

This line of reasoning would imply that we should take the CMs (and LMs, with the caveat above) in their default #1 senses, since we have no information to allow us to do anything else. Hence "Провал" should be taken in the context above to be Провал#1, its first sense, and so as a CM about failure not about an abyss, even thought the latter could conceivably be indicated by another context for the same words. This suggestion that the senses in a CM are major senses of the relevant words also implies that the two CMs above are different from each other, which preserves the insight of the tradition that metaphors are strictly speaking lies (attributed variously to Mark Twain, Nietzsche et al.) rather than the less acceptable alternative that CMs are tautologies, where the constituent senses simply recapitulate each other.

This risk of tautology in the expression of CMs is very real even if we are wary and assign (implicitly as main senses) interpretations to the symbols in CMs. If, in the CM [POVERTY IS A GAP], we allow the first WordNet sense interpretation to "gap" we get:

S: (n) **gap**, spread (a conspicuous disparity or difference as between two figures) "gap between income and outgo"; "the spread between lending and borrowing costs"

Thus, and depending on the sense assigned to "poverty", we have a very real risk of tautology since this sense of "gap" is itself abstract (and not, say, a gap between two pieces of wood) and itself very close to any definition of poverty, or at least "relative poverty" the currently fashionable version. This unfortunate fact can be dismissed, or simply accepted as a weakness or error in WordNet, or, perhaps, as a reason for excluding [POVERTY IS A GAP] as a CM.

One important inference from this discussion, if it has any value, is that we cannot just say, as many researchers in the Berkleyan universal metaphor tradition seem to want to, that some particular metaphor "in one language" is commoner than in another. As we have seen, it is a very sophisticated matter to establish whether LMs in two languages point to a single CM or not, given the problems of how any CM is to be unambiguously represented and, given the need for some lexical resource of at least the size and scope of (Euro)WordNet in order to do that. In the example above, the LM word strings in question in the two languages—Russian and English ---actually point to different CMs in the common interlingua, a conclusion that, we argued, undermines the foundation of the Berkeley approach to understanding metaphor, since the LMs could clearly be interpreted as "meaning the same thing". At this point, let us step

back and review the basic role of "preference" in detecting, then mapping, metaphors.

4. THE ROLE OF PREFERENCE IN DETECTING AND MATCHING METAPHORS

An exception to the "rule of main senses" we have just stated, as far as LMs are concerned, is the situation we have defined elsewhere as one of "conventional metaphor" [13] This is where a lexical resource such as WordNet actually encodes a metaphorical sense as a (dead or) conventional metaphor. Our approach to detecting metaphor has been that an initial *sufficient* criterion for a surface (LM) metaphor to be present is that a verb or adjective "preference" is broken [6] e.g. in the simplest case the verb does not receive the agent or object it expects (whether that last notion is unpacked linguistically or statistically) in a stereotypical case. Verbs and adjectives will, of course, have multiple senses in the lexicon, each with its own preferences. So to write *fall into poverty* is to break the preference for a spatial-container-like object for the basic sense of "fall into". This general criterion reappears frequently in the literature (e.g. the recent work of Shutova [4]) indeed it is not clear there is any alternative to it as a basic criterion for metaphor recognition, unless one believes that metaphors are detected by direct matching to stored CMs. As we have seen above this a notion whose very intelligibilitys dissolves somewhat under scrutiny.

If such preferences, and the associated noun-senses for fillers, are thought of as stored in a respository like WordNet or VerbNet, then what counts as a broken preference depends crucially on the state of lexicon at a given time, since sense inventories extend with time and indeed often come to store senses that were in origin metaphorical. Where that is the case, a dead, or as we would prefer to say conventional, metaphor will not result in a broken preference with respect to WordNet because in such a case the metaphorical sense is itself stored in WordNet and so will fit the demands of the corresponding verb.

So, to take a very simple and uncontentious example:

Public employees' unions have built a fortress around their pension systems

In VerbNet [14] we find the following:

[[VerbNet: build

   Member of

§build%2:31:03 (member of VN class base-97.1)

§build-26.1-1

•WordNet Sense 1

•Agent [+animate | +machine]

So "**Unions" violates Agent restriction for build**

•WordNet Sense 8

•Agent [+animate | +organization]

"Unions" satisfies the Agent restriction ---as an organization—for build]]

The situation is one where the primary sense of "build" is not satisfied by the first sense of the agent the sentence contains but is satisfied by a "lower' (in this case #8) sense. In [13] I proposed that this could serve as a useful heuristic (i.e. main sense failure but some lower sense a successful match) for detecting conventionalized metaphors of the sort this sentence contains, since such metaphors would be missed by any "preference breaking" heuristic for metaphor detection as there is a (lower) sense of "build" available for which the agent preference here is satisfied. The heuristic was that a main sense fails and a lower sense satisfies; and both parts must be true. Its main defect is that it relies on the ordering of senses in WordNet as carrying information, which is generally true but as always with this database has many errors and omissions.

The point here is not to draw attention to this metaphor detection heuristic against a large lexicon for its own sake, but only to show a limitation on the earlier suggestion that metaphor detection (and as we shall discuss below, metaphor mapping to CMs) must depend on the main senses, as listed in a lexicon. Our claim here is that this heuristic for detecting conventional or lexicalized metaphor does not compromise the general value of that rule. In the case of the above example, there are arguably two CM metaphors present: the major one is to do with barriers and the protection of assets, however expressed, and the other is more simply (and even though it is, more strictly, a meronym, though such differences are not crucial here):

ORGANIZATIONS ARE PEOPLE

which is expressed (in major senses of the relevant words) by the process of detection we have described.

The latter move is the basis of how preferences, and their violations in metaphor, are also central to the subsequent process of mapping from a detected metaphor to some stored form, which we are calling CMs. If we were again dealing with "He fell into poverty" we might expect the broken preference for the object of "fall into" to be some coding for hole/abyss/gap/aperture. The inference from that detection to the underlying metaphor in play is generally to assert that the metaphor's object (poverty in this case) is being asserted to be equivalent to the preferred filler that is made available in the lexical coding (e.g. in VerbNet, see [14]) but not in the sentence itself. This would lead directly to some form such as:

POVERTY IS AN ABYSS

as a potential CM, empirically derived from this example text rather than a linguist's intuition. The interesting difficulty is to determine at exactly what level its last term is to be expressed,

since "abyss" is, in general, a very magnified form of hole. The mapping process from a metaphor instance, or LM, to a CM, however expressed, will require an ontology of the kind that underlies WordNet to navigate from what appears in a VerbNet coding (perhaps "hole") to an item in an already stored CM (perhaps, as here, "abyss"). This method, merely sketched here, can in principle serve to map LMs to CMs, and to create potential CMs from text.

This process, making use of the preferred constituents of lexical codings, has been central to a number of systems based on inferences within lexical semantic structures and under names such as "projection" and "coercion" (e.g. Wilks, [6]; Pustejovsky,[15]; Nirenburg and Raskin, [16] and Hanks [17]) among many others. It provides at least the beginning of a process of determinate empirical construction of CMs from text cases quite different from the intuitive creation of CMs in the Berkeley tradition. Moreover, [22] contains a sophisticated analysis of some of the cross-lingual issues raised here. Further possible examples of the method would be with a failed subject+verb preference in *Israel has inflicted this wound on itself*. There we can get (from the stored VerbNet subject preference for "inflict" as PERSON) we can link the existing target (Israel) to the preferred subject (as source), namely PERSON, and then the WordNet type of "Israel" as COUNTRY to give as a possible CM: COUNTRY IS PERSON. We could do the same for verb+object failure as in: *The bank hyenas are feeding on money*, assuming we have access to "feed on" as a verb with its own preferences FOOD or EDIBLES. Then, using similar reasoning to that for subjects above, and again combining the assigned object and the preferred object, we can derive directly a potential CM: MONEY IS FOOD. For adjective+noun preferences, similar processes are possible, as in *Brazil's economic muscle will become increasingly important*. If we have a preference established for the preferred type of noun associated with the adjective "economic" as COMPLEX-SYSTEM, then from the existing adjective object "muscle" (and taking its semantic type from WordNet as BODY) we then have directly a CM: COMPLEX-SYSTEM IS BODY. Many metaphor theorists would want to argue that equations of target and source CMs produced by a process such as this must be brought under some higher level generalization on both sides of the assertion in the CM, as we shall now show.

Notice though that no claims here depend on the actual quality or completeness of resources such asVerbNet or WordNet. These are always variable, depending on the language used, and will always contain errors and omissions, as well as being constantly changing with the language itself. The only claim is that some such resource will be needed to carry out the processes described here, even if augmented in practice by statistical corpus computations (some of which augmented these resources in the work described in [13]).

There has been criticism of processes of this sort applied to the empirical construction of CMs in this manner: during a recent large-scale metaphor detection and interpretation project a project manager wrote:

"[CMs that were] proposed..... were inconsistent and generally unmotivated. For the most part, the relationship of an LM (for a Target) and a proposed CM was semantically extremely shallow with generally no mapping at all. This process caused a huge proliferation of "lexical" CMs, often dependent on a synset label from WordNet."[18]

It is odd, in the current empirical climate, to criticise a linguistic process for being grounded in data, rather than linguistic intuition. One must also respond (a) that there is no known correct *level* for the expression of CMs beyond the intuitions of metaphor theorists, so no level is demonstrably "too lexical" and (b) more fundamentally, the CMs are inevitably in some language (usually English) and require sense disambiguation of their terms, as we argued at length above. They are not in a language that is self-disambiguating, since nothing is. Hence the presence of WordNet labels, even if implicit, so as to indicate main senses as we suggested above, is inevitable. That would be a feature not a bug.

The problems of the appropriate level for the expression of CMs, their distance and separation from LMs and their very origins in intuition, are not ones that preoccupy only NLP researchers, as is clear from Deignan's:

".... at some points in the development of CMT [Conceptual Metaphor Theory], there has been a tendency for researchers to propose new conceptual metaphors using limited linguistic evidence. For instance, [19] take the idioms "he really couldn't swallow it" and "[leave] a bad taste in the mouth" as instantiations of a conceptual metaphor termed ACCEPTING SOMETHING IS EATING IT. It is not clear how many other realizations there might be of this conceptual metaphor, and in what way it differs from the more-often cited IDEAS ARE FOOD. Kovecses [20] lists as a conceptual metaphor CONSIDERING IS CHEWING, which again is difficult to separate from IDEAS ARE FOOD. If this tendency becomes widespread, the notion of a conceptual metaphor loses clarity, along with any predictive power it may have had." ([21] p.105)

I take the force of this comment, from a corpus linguistic standpoint, to be consistent with the NLP processing critique advanced in this paper, and indeed with the internal project critique quoted earlier above. However, there is a difference of emphasis here: Deignan argues that CMT theorists in fact make up CMs from data, no matter what they say about intuition, and I have argued that they should be constructed by a determinate process from data since there is no other reliable route. But the internal project critique earlier seems to say that derivation from data in any such way is a istake and leads to shallow CMs and "real" CMs come only from intuition. I hope I have set out reasons for thinking tis comment profoundly wrong and out of line with all modern thinking on linguistics and data.

5. THE LAKOFF BERKELEY VIEW OF METAPHOR REVISITED

This view, against which I have argued, seems to me to rest on the following, very questionable, assumptions:
1. There is a set of universal CMs, determinable by linguistic intuition and underlying all languages.

There is no suggestion this set should be small, even fixed, as Schankian primitives were once held to be, and certainly some

depend on developments in technology, economics etc. Yet, as I have argued, there is no empirical evidence for their existence or how many of them there are, and intuition as a source of linguistic insight is no longer considered reliable, taken alone. However, there may be a discovery procedure for them from text along the lines suggested here (and in [6]).

2. CMs can be expressed in an English-like language, whatever their real underlying representation.

I have argued that they are in fact in English, as they appear to be, and not as an inevitable approximation; this is made clear by the problem of expressing exactly what senses their constituent words are to be taken in. This situation is only tolerable as a heuristic if some form of cross-lingual sense representation is incorporated into the representation, as suggested here.

3. Surface metaphors (LMs) in languages can be mapped to these CMs in a determinate way.

I have argued that no definitive procedure is ever given, within this tradition, for performing this crucial step and it can only be attempted at all with the aid of some fairly reliable, cross-sense mapping of the languages concerned, such as (Euro)WordNet.

If LMs can be matched bottom-up to CMs in something like the way sketched here---as opposed to being the subject of some direct matching top-down from stored CMs to LMs in text---- it should be possible to count how many LMs correspond to a given CM. That would then make it possible to estimate the frequency of occurrence of CMs in a reliable manner. That analysis could be extended cross-lingually and cross-culturally if parallel text were available. Suppose we had an English-Spanish parallel text in which sentences are aligned. We could then ask whether LMs are detected in parallel (putatively synonymous) sentences and, if so, do they map to the same CMs. If they do, that would be independent confirmation of the utility or universality of such a CM. Quantitative and distributional questions about universal metaphor can only be asked, it seems to me, if procedures of this kind I sketch here are developed, but these are not obviously compatible with standard Lakoffian approaches to metaphor, though there is no reason in principle, or course, why it could not develop so as to incorporate some empirical theory of sense ambiguity like the present one.

My main conclusion is that, for these reasons, Berkeley metaphor theory cannot easily be the basis of an empirical exploration of metaphors in texts in multiple languages, and that any research program aimed at the interpretation and translation of metaphor instances so based will have been mistaken.

ACKNOWLEDGEMENTS

REFERENCES

[1] Lakoff, G., and Johnson, M. 1980 *Metaphors we live by*. University of Chicago Press: Chicago.

[2] Guo, C-M. 1989. *Constructing a machine-tractable dictionary from "longman dictionary of contemporary English*. Doctoral dissertation, Computer Science, New Mexico State University.

[3] Fass, D. and Wilks, Y. 1983. Preference semantics, ill-formedness and metaphor. *Journal of Computational Linguistics,* 9, pp.1798-187.

[4] Shutova, E., Teufel, S. and Korhonen, A. 2012. *Statistical Metaphor Processing*, Computational Linguistics, 39(2).

[5] Schank, R (Ed.,) 1975. *Conceptual Information Processing*. Elsevier: Amsterdam.

[6] Wilks, Y., 1968/2007. Making preferences more active. Reprinted in Ahmad, Brewster and Stevenson (Eds.) *Word and Intelligence I.* Springer: Berlin.

[7] Wilks, Y. 1977/2007. Good and bad arguments for semantic primitives. Reprinted in Ahmad, Brewster and Stevenson (Eds.) *Word and Intelligence I.* Springer: Berlin.

[8] Pulman, S. 1983. *Word meaning and belief*. Croom Helm: London.

[9] Lewis, D., 1972. General Semantics, In: Davidson and Harman (eds.), Semantics of natural language. Reidel: Dordrecht.

[10] Fodor, J. and Katz, J. 1963. The structure of a semantic theory. *Language*, *39*(2), Apr-Jun, 170-210.

[11] Montague, R. 1974. *Formal philosophy : selected papers of Richard Montague* / ed. and with an introd. by Richmond H. Thomason. New Haven: Yale Univ. Press.

[12] Vossen, P. (ed.) 1998. *EuroWordNet: a multilingual database with lexical semantic networks*. Kluwer: Amsterdam.

[13] Wilks, Y., Dalton, A., Allen., J.. Galescu, L. 2013. Automatic Metaphor Detection using Large-Scale Lexical Resources and Conventional Metaphor Extraction. *Proc.1st Workshop on Metaphor in NLP (Meta4NLP 2013)*. Atlanta, GA.

[14] Windisch Brown, S., Dligach, D., and Palmer, M. 2011 VerbNet Class Assignment as a WSD Task. In IWSC 2011:

Proceedings of the 9th International Conference on Computational Semantics, January 12 - 14, 2011, Oxford, UK.

[15] Pustejovsky, J. 1995. *The Generative Lexicon.* MIT Press: Cambridge, MA.

[16] Nirenburg, S. and Raskin, V. 2004. *Ontological semantics.* MIT Press: Cambridge, MA.
[17] Hanks, P. 2013. *Lexical analysis: norms and exploitations.* MIT Press: Cambridge, MA.

[18] iARPA:
http://www.iarpa.gov/Programs/ia/Metaphor/metaphor.html

[19] Gibbs, R., Bogdonovich, J., Sykes, J., & Barr, D. 1997. Metaphor in idiom comprehension. *Journal of Memory and Language*, 37, pp. 141-154.

[20] Kovecses, Z. 2002. *Metaphor: a practical introduction.* Oxford University Press: Oxford.

[21] Deignan, A. 2005. Metaphor and corpus linguistics. Benjamins: Amsterdam.

[22] Prior, A., Wintner, S., MacWhinney, B., & Lavie, A. (2011). Translation ambiguity in and out of context. *Applied Psycholinguistics, 32*, 93-111.

Relevance Theoretic Comprehension Procedures: Accounting for Metaphor and Malapropism

Zsófia Zvolenszky[1]

Abstract. According to Sperber and Wilson, relevance theory's comprehension/interpretation procedure for metaphorical utterances does not require details specific to metaphor (or nonliteral discourse); instead, the same type of comprehension procedure as that in place for literal utterances covers metaphors as well. One of Sperber and Wilson's central reasons for holding this is that metaphorical utterances occupy one end of a continuum that includes literal, loose and hyperbolic utterances with no sharp boundaries in between them. Call this the *continuum argument about interpreting metaphors*. My aim is to show that this continuum argument doesn't work. For if it were to work, it would have an unwanted consequence: it could be converted into a continuum argument about interpreting linguistic errors, including slips of the tongue, of which malaprops are a special case. In particular, based on the premise that the literal–loose–metaphorical continuum extends to malaprops also, we could conclude that the relevance theoretic comprehension procedure for malaprops does not require details specific to linguistic errors, that is, details beyond those already in place for interpreting literal utterances. Given that we have good reason to reject this conclusion, we also have good reason to rethink the conclusion of the continuum argument about interpreting metaphors.

1 INTRODUCTION

Mrs. Malaprop, a character in Sheridan's (1775) play *The Rivals* had a tendency to make linguistic errors of a special sort: she would describe people as being "the pineapple of politeness" (when she meant *pinnacle*); or "as headstrong as an allegory on the banks of the Nile" (when she meant *alligator*). Such slips of the tongue have since come to be called malaprops. In a framework like relevance theory, how might we characterize the process of interpreting malaprops as opposed to interpreting literal utterances? We will see that addressing this question exposes a challenge for the relevance theoretic treatment of *metaphorical* utterances.

Within philosophy of language as well as rhetoric the following claims are widely held, considered platitudinous even: the distinction between literal and figurative discourse carries theoretical importance, and metaphorical utterances clearly fall on the figurative side of the divide, constituting departures from

literality. Relevance theory calls into question these time-worn claims.

Relevance theory [1, 2] has become, over the past three decades, a leading research program in pragmatics. Its founders', Dan Sperber's and Deirdre Wilson's [3] most recent position on metaphorical utterances is that (i) the interpretation/comprehension procedure for metaphors does not require resources beyond those already needed to account for literal utterances (call this the *procedure claim*), and (ii) metaphorical utterances occupy one end of a continuum that includes literal, loose and hyperbolic utterances (call this the *continuum claim*). Relevance theorists seem to regard the continuum claim as one reason to hold the procedure claim; call this the *continuum argument about interpreting metaphors*.

Sperber and Wilson subscribe to this continuum argument: "We see this continuity of cases, and the absence of any criterion for distinguishing literal, loose, and metaphorical utterances, as evidence not just that there is some degree of fuzziness or overlap among distinct categories, but that there are no genuinely distinct categories, at least from a descriptive, psycholinguistic or pragmatic point of view. Even more important than the lack of clear boundaries is the fact *that the same inferential procedure is used in interpreting all these different types of utterance*" [3, p. 111–112, emphasis added].

In this paper, I aim to show that the continuum argument about metaphors, if it were to work, would face an unacceptable consequence: the argument would license *a continuum argument about interpreting malapropisms* (and more generally, a continuum argument about linguistic errors):

Continuum premise for malaprops: The literal–loose–metaphorical continuum extends to malaprops.

Procedure conclusion for malaprops: The relevance theoretic comprehension procedure for malaprops does not require details beyond those needed to account for literal utterances.

We have good reason to resist the malaprop conclusion: surely, when we manage to interpret Mrs. Malaprop as having meant 'alligator' when she said 'allegory', the fact that the lexically encoded meaning of 'allegory' becomes wholly irrelevant is a detail that is bound to be featured in a full description of our process of interpreting her. And if we want to resist the malaprop conclusion, then we have to find fault with the continuum argument about interpreting malapropisms. There are two strategies we could follow: we could fault the premise or fault the argument itself as non-truth-preserving. I will argue that the former strategy is not open to us, so our remaining option is to regard the malaprop argument as non-truth-preserving. But then we have to say the same about the continuum argument about interpreting metaphors also. Whether the comprehension

---

[1] Dept. of Logic, Institute of Philosophy, Eötvös Loránd Univ. (ELTE), Budapest, Hungary 1088. Email: `zvolenszky@elte.hu`.

[2] Our concern here is with acts if *linguistic* communication, but the communicative principle and the relevance theoretic framework are intended to apply to a broader range of cases: acts of ostensive communication which include, besides linguistic utterances, certain kinds of non-linguistic acts also.

procedure for interpreting metaphors includes any details specific to metaphor (or nonliteral discourse) therefore remains an open question.

2 RELEVANCE THEORY ABOUT THE LITERAL–LOOSE–METAPHORICAL CONTINUUM

Dan Sperber and Deirdre Wilson's influential framework for the study of communication, relevance theory [1, 2] outlines an inferential comprehension procedure that hearers follow in arriving at an interpretation of speakers' linguistic utterances. Crucially, the comprehension procedure is delimited and guided by specific assumptions about relevance (i)–(iii), accepted by speakers and hearers alike. (i) Cognition (generally, not just in the case of communication) aims to maximize relevance (this is the *cognitive principle* of relevance). (ii) Linguistic utterances communicate a presumption of their own optimal relevance (this follows from the *communicative principle* of relevance[2]). And (iii) an utterance is *presumed to be optimally relevant* if and only if it is at least relevant enough to be worth the speaker's effort to process it, and it is the most relevant utterance compatible with the speaker's abilities and preferences. The kind of inference involved in the relevance theoretic comprehension procedure is inference to the best explanation [4]. The concepts encoded by the words the speaker has used on a given occasion are mere starting points for arriving, via inferential steps, at an interpretation of her utterance: her utterance's explicit content (the speaker's explicit meaning) on the one hand, and its implicit content (which consists of implicit premises and conclusions) on the other.

By explicit and implicit content, we mean content that was *intended* as such by the speaker. The hearer's task is to *reconstruct* the explicit content and implicit premises and conclusions that the speaker has intended to communicate. Of course, rarely, if ever do hearers converge on the very same concepts as those that speakers actually meant. Nor is this required for successful communication. It suffices that the concepts reconstructed by the hearer be ones that allow him to draw (nearly enough) the same inferences as those intended by the speaker; it is enough that the reconstructed concepts "activate contextual implications that make the utterance relevant as expected" [3].

A recurring example of Sperber–Wilson's [3, 5, 6] exemplifies *loose use*: 'Holland is flat' uttered in the context of the following conversation: Peter and Mary are discussing their next cycling trip. Peter has just said that he feels rather unfit. Mary replies: "We could go to Holland. Holland is flat." Sperber–Wilson [5] illustrate the inferential comprehension procedure via which Peter interprets Mary's second sentence as follows.

(a) Mary has said to Peter, 'Holland is flat'.	*Decoding of Mary's utterance.*
(b) Mary's utterance is optimally relevant to Peter.	*Expectation raised by the recognition of Mary's utterance as a communicative act, and acceptance of the presumption of relevance it automatically conveys.*
(c) Mary's utterance will achieve relevance by giving reasons for her proposal to go cycling in Holland, which take account of Peter's immediately preceding complaint that he feels rather unfit.	*Expectation raised by (b), together with the fact that such reasons would be most relevant to Peter at this point.*
(d) Cycling on relatively flatter terrain which involves little or no climbing is less strenuous, and would be enjoyable in the circumstances.	*First assumption to occur to Peter which, together with other appropriate premises, might satisfy expectation (c). Accepted as an implicit premise of Mary's utterance.*
(e) **Holland is FLAT*** (where FLAT* is the meaning indicated by 'flat', and is such that Holland's being FLAT* is relevant-as-expected in the context).	*(Description of) the first enriched interpretation of Mary's utterance as decoded in (a) to occur to Peter which might combine with (d) to lead to the satisfaction of (c). **Interpretation accepted as Mary's explicit meaning.***
(f) Cycling in Holland would involve little or no climbing.	*Inferred from (d) and (e). Accepted as an implicit conclusion of Mary's utterance.*
(g) Cycling in Holland would be less strenuous, and would be enjoyable in the circumstances.	*Inferred from (d) and (f), satisfying (b) and (c) and accepted as an implicit conclusion of Mary's utterance.*

Table 1. Interpretation of Mary's utterance 'Holland is flat'.

As indicated on line (e) (in boldface), the explicit content of Mary's utterance 'Holland is flat' is 'Holland is FLAT*'. FLAT* is an *ad hoc concept* Peter arrived at that is distinct from, broader[3] than the lexicalized concept encoded by the word 'flat': say, FLAT. Unlike FLAT*, the extension of FLAT doesn't include imperfectly flat surfaces like the Dutch landscape.

Loose use, as in 'Holland is flat' is a type of literal discourse[4] that involves some departure from the lexically encoded concept. While the departure is greater than in many other instances of literal discourse, Sperber–Wilson [3] stress that the comprehension procedure for *some* literal utterances (to wit: cases of loose use) already involves the formation of *ad hoc* concepts. They suggest further that even in literal utterances that do not involve a departure from the lexically encoded concept, the process of disambiguating the expressions used involves inferential steps similar to those in Table 1. For example, Mary's and Peter's idiolect may have (at least) two senses associated with the word 'flat', one of which amounts to, say, "having a smooth, even surface" while the other, to "is in a horizontal position"; Sperber–Wilson [3, p. 111] suggest that if Mary uttered "My computer screen is flat", the process of interpreting her utterance and deciding that she has in mind the first and not

---

[3] Alternatively, according to another prominent relevance theorist, Robyn Carston [7], the formation of *ad hoc* concepts involves conceptual narrowing as well as broadening.
[4] Sperber–Wilson [1, pp. 234–235; 3] stress the literal status of instances of loose use.

the second sense of 'flat'would take a similar inferential procedure as the one seen in Table 1.

Sperber and Wilson [3] gradually build up a continuum a cases with no clear boundaries in between them. The continuum includes cases of disambiguation like ("My computer screen is flat"), various examples of

- loose use (or broadening), covering a broad range:
    - *Approximation*: "Holland is flat"';
    - *Limited category extension*: "Here is a Kleenex", said of a piece of non-Kleenex-brand tissue;
    - *Creative category extension*: "For luggage, pink is the new black";
- *Hyperbole*: "Joan is the kindest person on earth";
- *Nonpoetic metaphor*: "Joan is an angel";
- *Poetic metaphor*: "The fog comes on little cat feet" (from Carl Sandburg's poem *The Fog*).

A central claim of relevance theory (besides Sperber and Wilson, see also Carston [7]) is that each of the listed cases involves the formation of an *ad hoc* concept, one that—as we go down the list of examples—exhibits a gradually greater degree of departure from the concept lexically encoded by the word used, that is, the concept that serves as one of the starting points for the comprehension procedure. The *ad hoc* concepts are then featured as part of the explicit content attributed to the speaker (as in line (e) in Table 1). The *ad hoc* concepts for the listed examples (except for poetic metaphors, to be discussed in detail in Section 4) are as follows:

- FLAT*, whose extension includes imperfectly flat surfaces like the Dutch landscape;
- KLEENEX*, whose extension includes paper tissues that aren't Kleenex brand;
- BLACK*, whose extension includes (roughly) objects of a fashionable, trendy color, among them pink suitcases;
- KINDEST PERSON ON EARTH*, whose extension includes people who are very kind, but not even close to being among the *kindest*;
- ANGEL*, whose extension includes nonangelic human beings who are very kind.

We are now in a position to formulate in far more depth and detail Sperber–Wilson's (and other relevance theorists') argument about interpreting metaphors:

THE CONTINUUM ARGUMENT ABOUT INTERPRETING METAPHORS

Continuum premise for metaphors:
All metaphorical utterances (poetic and nonpoetic alike) can be located on a continuum of cases that includes loose use (a kind of literal use) as well as hyperbolical and metaphorical uses. Further, the process of forming *ad hoc* concepts to arrive at the explicit content attributed to the speaker is a tool that is readily applicable to all metaphorical utterances (not just to instances of loose use and hyperbole).

Procedure conclusion for metaphors:
Equipped with the relevance theoretic comprehension procedure and the *ad hoc* concept formation tool, both already required for interpreting literal utterances like loose use, we have all the resources needed to describe the comprehension procedure at play during the interpretation of metaphorical utterances. No further details specific to metaphor (or figurative language use) are needed in a comprehensive account of interpreting metaphors.

In Section 3, I will raise an objection intended to show that the continuum argument about interpreting metaphors is flawed: even if we accepted its premise, that is not reason enough to accept its conclusion also. I will motivate this by giving what I think is an analogous argument about malaprops with a clearly false conclusion. Someone might then raise a counterobjection: the argument about malaprops has a false conclusion because its premise is false. So as long as we can maintain (as relevance theorists do) the continuum premise for metaphors while resisting its analog about malaprops, we are entitled to keep the continuum argument about interpreting metaphors and maintain that its conclusion is true because its premise is. In Section 4, I will elaborate this counterobjection and deflect it by showing that the malaprop premise and the metaphor premise are equally plausible. My objection therefore has traction and there is room to reject the procedure conclusion for metaphors, despite relevance theorists' arguments to the contrary.

3 AN OBJECTION TO THE CONTINUUM ARGUMENT ABOUT INTERPRETING METAPHORS

Once we have accepted the continuum argument about interpreting metaphors, along with its premise and its conclusion, we have, I claim, no reason to resist making the same moves with respect to an analogous argument about malaprops (and more generally, about linguistic errors):

THE CONTINUUM ARGUMENT ABOUT INTERPRETING MALAPROPS

Continuum premise for malaprops:
All malaprops can be located on a continuum of cases that includes loose use (a kind of literal use) as well as hyperbolical and metaphorical uses. Further, the process of forming *ad hoc* concepts to arrive at the explicit content attributed to the speaker is a tool that is readily applicable to all malaprops.

Procedure conclusion for malaprops:
Equipped with the relevance theoretic comprehension procedure and the *ad hoc* concept formation tool, both already required for interpreting literal utterances like loose use, we have all the resources needed to describe the comprehension procedure at play during the interpretation of malaprops. No further details specific to slips of the tongue (or more broadly: linguistic errors) are needed in a comprehensive account of interpreting malaprops.

But—my objection goes—there is a flaw in this argument: (i) its conclusion is clearly unacceptable and (ii) it remains unacceptable even if we accept its premise. And if we accept all this, we have exposed a flaw in the original continuum argument about interpreting *metaphors*. In the rest of this section, I aim to establish (i), in the next section, (ii).

The procedure conclusion for malaprops leads to the following bizarre results:

- *Allegory example.* In interpreting Mrs. Malaprop's utterance "She is as headstrong as an allegory on the banks of the Nile", the explicit content that hearers arrive at involves an *ad hoc* concept ALLEGORY*, which is constructed by broadening the concept lexically encoded by the word 'allegory' (about a certain kind of trope or figure of speech) in such a way that its extension includes *alligators*. The comprehension procedure is basically the same as that in Table 1, it's just that the degree of departure to get from FLAT to FLAT* is not as great as that from ALLEGORY to ALLEGORY*.
- *Spanking example.* In interpreting George W. Bush's utterance in the context of a speech he gave at a school "I want to spank all teachers" (he meant *thank all teachers*), the explicit content that hearers arrive at involves an *ad hoc* concept SPANK*, which is constructed by broadening the concept lexically encoded by the word 'spank' (about slapping) in such a way that its extension includes acts of *thanking*. The comprehension procedure is basically the same as that in Table 1, it's just that the degree of departure to get from FLAT to FLAT* is not as great as that from SPANK to SPANK*.

As mentioned before, the continuum argument about malaprops is readily extended to linguistic errors of all sorts, including slips of the tongue other than malaprops as well as mistaken translations like the following:

- *Steak example.* In interpreting a German speaker's order in a restaurant "I want to become a steak" ('bekommen' in German means 'get'), the explicit content that hearers arrive at involves the *ad hoc* concept BECOME*, which is constructed by broadening the concept lexically encoded by the word 'become' in English (about 'turning into') in such a way that its extension includes one thing *getting* another. The comprehension procedure is basically the same as that in Table 1, it's just that the degree of departure to get from FLAT to FLAT* is not as great as that from BECOME to BECOME*.

It is bizarre to think that when we manage to interpret successfully the German speaker's request to "become a steak", we are broadening the concept lexically encoded by the English word 'become'. After all, our grasping that he's talking about getting a steak rather than turning into one happens *despite* his use of the English word 'become'. We can say the same about understanding Mrs. Malaprop's and George W. Bush's utterances: it is *despite* the encoded meaning of the words they have used that we manage to interpret them as having said something about alligators and thanking, respectively.

In the light of this, it seems that relevance theoretic comprehension procedures, as they stand, are missing key details that distinguish malaprops (and more broadly, linguistic errors) from utterances that are literal or metaphorical. To wit: the procedure has to specify that in utterances like 'Holland is flat', 'Joan is an angel' (loose and metaphorical uses alike), the speaker has *not* committed a linguistic error; further, that the speaker (and hearer) takes the lexically encoded concept

associated with her words to be in force, and would not retract her words when confronted with the concept lexically encoded by her words. By contrast, in the case of linguistic errors including malaprops, the hearer is rerouting the inference such that he sets aside the lexically encoded concept entirely, and the speaker, when confronted with the lexically encoded concept, would retract his or her words: "I didn't mean spanking teachers was desirable, I wanted to talk about thanking them." "I didn't mean there were allegories on the banks of the Nile, I wanted to talk about alligators". But we would have absolutely no grounds for seeking such additional details if we thought the continuum argument about malaprops worked and moreover featured a true premise. If, despite the argument about malaprops, we thought the additional details were needed, then we open the door to seeking additional details with which to supplement the comprehension procedure for metaphorical utterances also. And we thereby open the door to rejecting the conclusion of the continuum argument about metaphors.

An analogy helps illuminate what my objection, if successful, shows with respect to Sperber–Wilson's continuum argument about interpreting metaphors. If you are at Columbus Circle in Manhattan and want to take the subway to the Museum of Natural History (at 81st Street), then don't get on the A train (the 8th Avenue Express); despite the fact that you would initially approach your desired destination, eventually, your train would whizz right past the Museum of Natural History, taking you all the way to 125th Street in Harlem, far away from your desired destination. Likewise: if you don't want an inferential comprehension procedure for malaprops (and other linguistic errors) that invokes no more than the formation of *ad hoc* concepts at work in the comprehension procedure you posited for cases of loose use, then don't apply the continuum argument to metaphorical utterances, for you won't be able to get off there but will be whisked straight to a place where you don't want to be: the continuum argument about interpreting malaprops.

4 A COUNTEROBJECTION DEFLECTED

It seems natural to respond to the foregoing objection as follows: a distinguishing feature of linguistic errors, malaprops included, is that the speaker makes a mistake about which *word form* is associated with the lexically encoded concept that he or she wants to express: G. W. Bush has said 'spank' even though his intended concept is expressed by the word form 'thank'; Mrs. Malaprop has said 'allegory' even though her intended concept is expressed by the word form 'alligator'. Those voicing such a counterobjection may then continue: of course the swapping of word forms, and the fact that the hearer recognizes the swap and reroutes the inference accordingly, will be part of the comprehension procedure via which he interprets malaprops and the like. We are in no way forced to regard the alligator, spanking and steak examples as cases involving simply the formation of *ad hoc* concepts with extreme degrees of departure from the lexically encoded concepts that had served as starting points for the construction of the *ad hoc* concept. This is how the counterobjection goes.

Someone could maintain this line while holding on to the continuum argument for *metaphors* and its conclusion, by denying the premise of the continuum argument about *malaprops*. This would amount to showing either that—in the context of relevance theory—extending the literal–metaphorical

continuum to malaprops (and other linguistic errors) is unfounded, or that—again, in the context of relevance theory—extending the tool of *ad hoc* concept construction to malaprops (and other linguistic errors) is unfounded. In what follows, I will show that neither of these will work and hence the counterobjection fails. My response consists of three parts:

(A) In the case of poetic metaphors, the *ad hoc* concept departs greatly from the lexically encoded one, yet Sperber–Wilson (and others) do not doubt that here, too, explicit content is arrived at via the construction of an *ad hoc* concept.

(B) With respect to malaprops (and other linguistic errors also) we can talk about a continuum of cases ranging from limited to extreme degrees of discrepancy between the intended concept and the lexically encoded one. And the limited-discrepancy cases fit squarely the *ad hoc* concept formation paradigm and can be readily placed on the literal–metaphorical continuum Sperber–Wilson had posited.

(C) In formulating the continuum argument about metaphors, Sperber–Wilson appealed to considerations (about there being a continuum of cases that encompasses various types of loose use, including approximation, limited and creative category extension, along with hyperbole, nonpoetic metaphor and poetic metaphor) based on which there is no reason to deny that the continuum and the process of *ad hoc* concept formation extends to all other examples that (i) themselves form a continuum, (ii) are candidates for being accounted for via the already posited inferential comprehension procedure featuring the formation of *ad hoc* concepts, and (iii) include clear candidates for inclusion on the literal–metaphorical continuum.

In Section 2, I have already given reasons for holding (C). In what follows, I will, in turn, motivate (A) and (B).

(A) concerns poetic metaphors. We've already encountered the example from Sandburg's poem "The fog comes on little cat feet". According to Sperber–Wilson, the explicit content arrived at in the comprehension procedure for interpreting this line of the poem involves the *ad hoc* concept: ON-LITTLE-CAT-FEET*. What Sperber–Wilson say about this concept signifies that it involves a great degree of departure from the lexically encoded concept: the *ad hoc* concept is supposed to help convey that the fog is spreading in a smooth, quiet, stealthy and deliberate way. Yet it remains quite vague what this *ad hoc* concept is, in what direction it takes off from the lexicalized concept, what does and does not belong in its extension. The authors offer us limited guidance on these matters: ON-LITTLE-CAT-FEET* "is the concept of a property that is difficult or impossible to define, a property possessed in particular by some typical movements of cats (though not all of them—little cat feet can also move in violent or playful ways) and, according to the poem, by the fog" [3, p. 122].

As Sperber–Wilson see it, the great distance between lexicalized and *ad hoc* concepts and the vague description of the latter is no obstacle to applying the *ad hoc* concept formation paradigm to highly creative, poetic metaphors. Then comparably great distances and vagueness characterizing ALLEGORY* (whose extension includes certain reptiles) and SPANK* (whose extension includes acts of thanking) should be no obstacle to applying the

ad hoc concept formation paradigm to malaprops (and other linguistic errors).

Turning to (B), about examples involving limited-discrepancy between the encoded concept and the intended one. Examples like the following form a continuum with the extreme-discrepancy examples about allegory, spanking and becoming a steak. Meanwhile, these examples fit squarely within the *ad hoc* concept formation paradigm, comparable to the "Here is a Kleenex" and "For luggage, pink is the new black" type examples.

Ocean example (a slip of the tongue involving limited discrepancy). G. W. Bush said once: "I didn't grow up in the ocean—as a matter of fact—near the ocean—I grew up in the desert. Therefore, it was a pleasant contrast to see the ocean. And I particularly like it when I'm fishing." In interpreting the first portion of Bush's utterance, via *ad hoc* concept formation, from the encoded lexical meaning IN-THE-OCEAN, we arrive, by broadening, to IN-THE-OCEAN*, whose extension includes events and things *near* the ocean.

Library example (a mistaken translation involving limited discrepancy). A French speaker says: "There is a library around the corner" to mean that there is *bookshop* around the corner (in French 'libraire' means bookshop). In interpreting the utterance, via *ad hoc* concept formation, from the encoded lexical meaning of LIBRARY, we arrive, by broadening, to LIBRARY*, whose extension includes bookshops. (Such an utterance could also exemplify a slip of the tongue involving limited disrepancy.)

In the ocean example, the distance between IN-THE-OCEAN and IN-THE-OCEAN* is no greater and no less vaguely delineated than that between KLEENEX and KLEENEX*. The same can be said about LIBRARY and LIBRARY* also. And we can envision a continuity of cases from such limited-discrepancy examples to the more extreme ones like in the allegory, spanking and steak examples.

This concludes my justification for (A)–(C), which together show that the counterobjection about swapped word forms does not undermine the objection I had formulated against the continuity argument about interpreting metaphorical utterances. After all, the limited-discrepancy examples of linguistic error make clear that the continuum premise for malaprops (and other linguistic mistakes) is just as plausible as the continuum premise for metaphors. We therefore have at hand two analogous arguments, both with true premises, and the one about malaprops boasting a clearly false conclusion. Hence, the other argument, about metaphors, is also undermined: the truth of its premise is no guarantee for the truth of its conclusion.

5 CONCLUSION AND FUTURE WORK

The continuum argument about interpreting metaphorical utterances is central to Sperber–Wilson's conclusion that "[t]here is no mechanism specific to metaphors, no interesting generalisation that applies only to them. In other terms, linguistic metaphors are not a natural kind, and 'metaphor' is not a theoretically important notion in the study of verbal communication" [3, p. 97]. My aim has been to show that we need not accept this conclusion given that the continuum

argument about interpreting metaphors is flawed, as shown by its application to malaprops (and other linguistic errors).

In the wake of my objection to the continuum argument, several questions arise.

First, what shall we make of empirical considerations about metaphor processing, according to which, for example, the interpretation procedure for simpler metaphors is similar to that for literal utterances, while interpreting highly creative or novel metaphors involves a markedly different procedure [8]?[5] The dialectical situation is as follows: such considerations support or undermine, *independently of the continuum argument about interpreting metaphors,* the claim that a similar comprehension procedure applies to literal utterances and certain types of metaphorical utterances. The continuum argument doesn't—cannot—provide an objection to or further support for such claims, because (as I have tried to argue, successfully, I hope) if it were to work, it would show too much, so it doesn't work. Therefore the tenability of the claim about a literal–loose–metaphorical continuum and the application of *ad hoc* concept formation in the interpretation of metaphorical utterances will depend on *other* (experimental-data-driven) arguments.

Second, it is worth considering a positive proposal about how to supplement the relevance theoretic comprehension procedure for interpreting metaphors. I address this question in work in progress [11, 12], drawing in part on some of the considerations that provide missing details to supplement the comprehension procedure for interpreting malaprops and other linguistic errors (these were briefly discussed in Section 3). In the case of metaphorical utterances (but not malaprops), the speaker (and hearer) takes the lexically encoded concept associated with her words to be in force, and would not retract her words when confronted with the concept lexically encoded by her words. "The fog doesn't really walk on feline legs," someone might challenge the poet. And he might reply: "I was speaking metaphorically. But I stand by my words: The fog does come on little cat feet". By contrast, Mrs. Malaprop, when challenged, "There are no such things as pineapples of politeness," would (likely) respond: "I retract my previous words; I meant to speak about a *pinnacle* of politeness".[6] Such differences in the response to being challenged about the lexically encoded concepts associated with one's words do, I think, offer a promising starting point for the sorts of details that a relevance theoretic comprehension procedure might incorporate in an account of metaphor. Such an account would part ways with Sperber–Wilson's stance, claiming instead that there are, after all, interesting details and generalizations specific to metaphors. More generally, the various ways in which lexically encoded concepts systematically constrain speakers' meaning in the case of loose, hyperbolic and metaphorical utterances is a worthy area of inquiry within the relevance theoretic framework.[7]

---

[5] More recent experimental results [10] cast doubt on earlier views positing a marked difference in the processing of novel metaphors and literal utterances. Carston [9] a central figure of relevance theory parts ways with Sperber–Wilson [3] and posits two distinct modes of processing metaphorical utterances.

[6] See Camp [13, 14] about how deniability reveals distinctive features of metaphorical utterances.

[7] I have received many incisive comments in connection with this research project. I thank audiences and organizers at two conferences (Philosophy of Linguistics and Language X, in the Special Session on Dan Sperber és Deirdre Wilson's Philosophy of Language,

REFERENCES

[1] D. Sperber and D. Wilson. *Relevance: Communication and Cognition*, Blackwell, Oxford, UK (1986) (2nd edition 1995).

[2] D. Wilson and D. Sperber. *Meaning and Relevance*, Cambridge University Press, Cambridge, UK (2012).

[3] Sperber, Dan – Wilson, Deirdre. A deflationary account of metaphors. In: *The Cambridge Handbook of Metaphor and Thought.* R.W. Gibbs (Ed.).Cambridge University Press, Cambridge, UK 84–105 (2008) (reprinted in Wilson–Sperber 2012).

[4] N. Allott. Relevance Theory. In: *Perspectives on Linguistic Pragmatics* (*Perspectives in Pragmatics, Philosophy & Psychology 2*). A. Capone, F.L. Piparo, M. Carapezza (Eds.). Springer, Berlin, Germany 57–98 (2013).

[5] D. Wilson and D. Sperber. Truthfulness and Relevance. *Mind* 111:583–632, 2002 (reprinted in Wilson–Sperber 2012).

[6] D. Sperber and D. Wilson. Pragmatics. In: *Oxford Handbook of Contemporary Analytic Philosophy*. F. Jackson, M. Smith (Eds.), Oxford University Press, New York, NY, US 468–501 (2005) (reprinted in Wilson–Sperber 2012).

[7] R. Carston. *Thoughts and Utterances*: *The Pragmatics of Explicit Communication*, Blackwell, Oxford, UK (2002).

[8] R.W. Gibbs, *The Poetics of Mind: Figurative Thought, Language and Understanding*, Cambridge University Press, Cambridge, UK (1994).

[9] R. Carston. Metaphor: Ad Hoc Concepts, Literal Meaning and Mental Images. *Proceedings of the Aristotelian Society* 110:295–321.

[10] B. Forgács, Á. Lukács and C. Pléh. Lateralized Processing of Novel Metaphors: Disentangling Figurativeness and Novelty. *Neuropsychologia* 56:101–109 (2014) doi:10.1016/j.neuropsychologia.2014.01.003

[11] Z. Zvolenszky. Inferring Content: Metaphor and Malapropism. Manuscript, Department of Logic, Institute of Philosophy, Eötvös Loránd University (ELTE), Budapest, Hungary (2015).

[12] Z. Zvolenszky. Why Inflate Sperber & Wilson's Account of Metaphor? Manuscript, Department of Logic, Institute of Philosophy, Eötvös Loránd University (ELTE), Budapest, Hungary (2015).

[13] E. Camp. Showing, Telling and Seeing. *The Baltic Yearbook of Cognition, Logic and Communication*, vol. 3, *Figure of Speech*, 1–24 (2008).

[14] E. Camp. Sarcasm, Pretense, and The Semantics/Pragmatics Distinction. *Noûs* 46(4):587–634 (2012). doi: 10.1111/j.1468-0068.2010.00822.x