Two Analogy-Strategies:

The Cases of Mind Metaphors and Introspection

AUTHOR: Eugen Fischer, School of Politics, Philosophy, Language and Communication Studies, University of East Anglia, Norwich NR4 7TJ, UK.

Tel.: +44 (0)1603 593416. Email: e.fischer@uea.ac.uk

ABSTRACT: Analogical reasoning is often employed in problem-solving and metaphor interpretation. This paper submits that, as a default, analogical reasoning addressing these different tasks employs different mapping strategies: In problem-solving, it employs analogy-maximising strategies (like structure mapping, Gentner & Markman 1997); in metaphor interpretation, analogy-minimising strategies (like ATT-Meta, Barnden 2015). The two strategies interact in analogical reasoning with conceptual metaphors. This interaction leads to predictable fallacies. The paper supports these hypotheses through case-studies on 'mind'-metaphors from ordinary discourse, and abstract problem-solving in the philosophy of mind, respectively: It shows that (1) default metaphorical interpretations for vision- and space-cognition metaphors can be derived with a variant of the analogy-minimising ATT-Meta approach, (2) philosophically influential introspective conceptions of the mind can be derived with conceptual metaphors only through an analogy-maximising strategy, and (3) the interaction of these strategies leads to hitherto unrecognised fallacies in analogical reasoning with metaphors. This yields a debunking explanation of introspective conceptions.

KEYWORDS: Analogical inference, conceptual metaphor, metaphor interpretation, mind metaphors, introspection, debunking explanation

ACKNOWLEDGMENTS: For helpful comments on previous drafts and closely related material I am indebted to John Barnden, Christian Burgers, and Rachel Giora, an anonymous referee for this journal, and audiences at the 8th AISB Symposium on Computing and Philosophy, Canterbury, UK, April 2015, the 11th Conference of the Association for Researching and Applying Metaphor, Berlin, Germany, July 2016, and 6th conference of the Scandinavian Association for Language and Cognition, Lund, Sweden, April 2017. This research was funded through QR funding by the University of East Anglia.

I am unaware of any relevant financial interest or benefit.

LENGTH: Text (incl. footnotes and tables) 14,560. References: 2,099.

Two Analogy-Strategies:
The Cases of Mind Metaphors and Introspection

ABSTRACT: *Analogical reasoning is often employed in problem-solving and metaphor interpretation. This paper submits that, as a default, analogical reasoning addressing these different tasks employs different mapping strategies: In problem-solving, it employs analogy-maximising strategies (like structure mapping, Gentner & Markman 1997); in metaphor interpretation, analogy-minimising strategies (like ATT-Meta, Barnden 2015). The two strategies interact in analogical reasoning with conceptual metaphors. This interaction leads to predictable fallacies. The paper supports these hypotheses through case-studies on 'mind'-metaphors from ordinary discourse, and abstract problem-solving in the philosophy of mind, respectively: It shows that (1) default metaphorical interpretations for vision- and space-cognition metaphors can be derived with a variant of the analogy-minimising ATT-Meta approach, (2) philosophically influential introspective conceptions of the mind can be derived with conceptual metaphors only through an analogy-maximising strategy, and (3) the interaction of these strategies leads to hitherto unrecognised fallacies in analogical reasoning with metaphors. This yields a debunking explanation of introspective conceptions.*

1. Introduction

Analogical reasoning is an engine of creative thought and language use: Its use in problem-solving has been studied in artificial intelligence (review: Gentner & Forbus 2011), cognitive psychology (review: Holyoak 2012) and the philosophy of science (review: Bartha 2013); its use in motivating and interpreting metaphorical expressions is a central tenet of Conceptual Metaphor Theory (Lakoff & Johnson 1980, 1999, extension: Steen 2011, review: Gibbs 2011), has been studied in artificial intelligence (review: Barnden 2008), and has experimental support from cognitive psychology (Gentner et al. 2001, Bowdle & Gentner 2005).

This paper will put forward and philosophically deploy the new *differential processing hypothesis* that analogical reasoning employs different mapping-strategies in problem-solving and metaphor-interpretation, respectively: As a default, our hypothesis claims, analogical reasoning in problem-solving employs analogy-maximising mapping-strategies; as a default, what analogical reasoning is involved in metaphor-interpretation uses restricted or analogy-minimising mapping-strategies. I.e.: When using analogies in problem-solving, we try to maximise the correlations between source-model and target and row back only where this leads to absurd conclusions; in metaphor interpretation, we try to minimise those correlations, and add new ones only in rare cases where we otherwise fail to make sense of people's talk.

Most computationally-implemented models of analogy follow analogy-maximizing strategies (Hodgetts et al. 2009), the best-known models being structure mapping theory SMT (Gentner 1983, Gentner & Bowdle 2008) (implemented as Structure Mapping Engine SME, Falkenhainer et al. 1989, Forbus et al. 2016), ACME (Holyoak & Thagard 1989) and LISA (Hummel & Holyoak 1997). By contrast, there are only three detailed computational models of restrictive mapping and inference strategies: Barnden's ATT-Meta (Barnden 2008, 2015, Lee & Barnden 2001), Hobbs's (1992) and Narayanan's (1999) models. The use of analogy-

maximising strategies in problem-solving is widely assumed. The influential SMT theory has extended this approach to metaphor interpretation (Gentner & Bowdle 2008, Wolff & Gentner 2011). Here, it competes directly with analogy-minimising approaches: In particular, the ATT-Meta model has been developed with a view to capturing analogical reasoning in metaphor interpretation, and supported by showing that reasoning with restricted mappings delivers accurate interpretations for a wide range of examples (Barnden 2001, Lee & Barnden 2001)[1] and can elegantly model linguistic phenomena including mixing of metaphors (Barnden 2016) and the open-endedness of extended metaphors (Barnden & Lee 2001, Barnden 2015).

On this basis, the present paper will argue for the differential processing hypothesis through a case study that simultaneously brings out the importance of distinguishing between the two mapping-strategies, namely, by showing how their interaction in analogical reasoning with metaphors leads to fallacies. Experimental studies from cognitive and social psychology as well as communication science suggest such reasoning spontaneously occurs in problem-solving (Thibodeau & Boroditsky 2011, 2013, 2015) under conditions characteristic of much philosophical thought: high level of abstraction (Keefer et al 2014), greater psychological distance (Jia & Smith 2013), low confidence in own target-domain understanding (Landau et al 2014), and low level of target-domain knowledge (Vandeleene et al, in prep). Indeed, various philosophers have suggested that such reasoning is at the root of philosophically and culturally influential introspective conceptions of the mind (Fischer 2011, Lakoff & Johnson 1980, 1999, Rorty 1980). We therefore proceed from a case-study on 'mind'-metaphors in ordinary discourse and on introspective conceptions of the mind: We will show that the analogy-minimising approach can account for metaphorical default interpretations of the ordinary talk, while the analogy-maximising approach is needed to explain introspective conceptions, as formulated in abstract philosophical problem solving.

We will first explain the restricted mapping strategy of ATT-Meta (Section 2) and illustrate how the model works by using it to interpret vision cognition metaphors (Section 3). Second, we will develop the AI-based approach further through integration with key findings from psycholinguistics (Section 4) and show how the resulting 'minimal analogy theory' of extended metaphor can account for spatial cognition metaphors that are the home of 'the mind' in ordinary discourse (Section 5). Third, we will show that the central tenets of the targeted philosophical conceptions of the mind can be obtained – only – through analogy-maximising reasoning (Section 6), and use these analyses of ordinary mind-talk and philosophical conceptions, respectively, to expose two fallacies – one local, one systematic, both frequently made – in the analogical reasoning with metaphors that underlies the introspective conceptions (Section 7).

Reconstructing different analogy strategies explored by AI research allows us to identify philosophically relevant fallacies that have not been recognised previously. Crucially, it allows us to do so in the current (and arguably not contingent) absence of comprehensive normative theories of analogical inference (Bartha 2010, 2013). This will allow us to contribute to the development of a (cooperative) naturalised 'cognitive epistemology' (Fischer et al. 2015) that

[1] These studies analyse examples from the Berkeley Master Metaphor List (Lakoff 1994) and Goatly (1997).

shows us when and where thinkers may (not) go along with heuristic inferences that strike them as plausible – a key aim of the 'Sources Project' emerging from experimental philosophy (Fischer & Engelhardt 2016, Weinberg 2015).

2. Analogies for Metaphor: The ATT-Meta Model

As standardly conceived in the overlapping fields of artificial intelligence (review: Gentner & Forbus 2011) and cognitive psychology (review: Holyoak 2012), 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.). This step is governed by semantic and structural constraints: According to the arguably most influential analogy-maximising model of analogical inference (SMT), we first correlate SD and TD 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) (Markman 1997, Markman & Gentner 2005). 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 (Holyoak 2012): 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]
(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 (entity or relation) unchanged into the conclusions ('*generation*').

We will consider philosophically pertinent examples below (Section 6).

Conceptual Metaphor Theory assumes that analogical reasoning is involved in initially motivating the metaphorical extension of whole families of related expressions and may subsequently be employed in interpreting expressions belonging to such extended metaphors, especially when speaker/hearers first encounter them. A case in point is the extended metaphor KNOWING AS SEEING:

> 'It is *clear* or *obscure* to me why you did what you did, according to whether or not I manage to *see* any reasons for acting that way. I may *look for* reasons where these are *hidden* or *be blind to* reasons *in plain view*. An *illuminating* explanation which *throws new light* on your action will let me *discern* reasons I had previously *overlooked*, and thus *get a fuller picture* of these reasons, where I was previously *completely in the dark*.'

[2] Throughout, properties or 'object-attributes' are here treated as 1-place relations.

According to standard versions of the theory (e.g., Lakoff & Johnson 1980, 1999), pertinent analogical reasoning employs *conceptual metaphors*, i.e., comprehensive source-target mappings which preserve relations and may be recruited for interpreting and reasoning with metaphors. On a standard account, they map, e.g., source-domain concepts like 'visually clear', 'seeing', and 'visually focusing on', etc., onto target-domain concepts like 'intellectually clear', 'knowing', and 'mentally focusing on', etc. I will call such mappings *wide conceptual metaphors* when it is possible to generate them through an analogy-maximising mapping strategy like that of SMT (above).

In a seminal paper, Grady (1997) deconstructed such wide conceptual metaphors into mixtures of more restricted bundles of mappings ('primary metaphors'). The ATT-Meta model (Barnden 2008, 2015, Barnden et al. 1997, Lee & Barnden 2001) economises yet further on source-target mappings, and makes the most out a limited stock of familiar mappings, complemented by even fewer widely applicable mappings: For the kind of analogical reasoning potentially involved in interpreting metaphors like those sampled, the computationally implemented model makes do with *'narrow conceptual metaphors'* (aka 'metaphorical views') which it systematically unfolds from *core mappings or correspondences* like

(1) S sees X ↔ S knows what X is (cp. 'I see your point.')

(2) S looks at X ↔ S thinks about X (cp. 'Let's look at the issue more carefully.')

To derive metaphorical interpretations of utterances, the computational model deploys mainly core mappings which map relations (rather than properties or objects), lexicalised by verbs or verb phrases,[3] and generally prefers mappings at higher levels of abstraction to mappings at lower levels. While the model remains silent on the origin of these core mappings, it is consistent with different explanations of how such mappings come to be made, and with the use of different explanations for different core mappings, ranging from pragmatic strengthening (Traugott 1989) of stereotypical inferences (which may account, e.g., for core correspondence (1)) to perceptual simulation theory (Barsalou 2007, Gibbs 2006).

Further mappings are built up from and around these cores: Where generic or domain-neutral functions, properties, or relations qualify relations etc. that already get mapped, they are, as a default, carried over by *generic mapping adjuncts* which apply to correspondences regardless of the domains they link. The simplest such adjunct deals with the logical function of negation:

(NEG) IF a relation Rxy [e.g. x looks at y] in the SD
CORRESPONDS TO
a relation R*xy [e.g. x thinks about y] in the TD
THEN the relation not-Rxy [e.g. x does not look at y] in the SD
CORRESPONDS TO
the relation not-R*xy [e.g. x does not think about y] in the TD.

[3] This focus is sometimes obscured by the traditional THING1 AS THING2 labels which the ATT-Meta literature continues to apply to mappings that actually correlate relations. Thus, e.g., MIND AS PHYSICAL SPACE actually correlates the relations 'J is physically located in a physical region belonging to [person] P' and '[Person] P is able mentally to use [idea] J' (e.g., Barnden 2016).

This adjunct generates a new correspondence for any correspondence it receives as an input. Similar mapping adjuncts deal with ability and attempts to V-y, inclinations to V-y, etc. E.g.:

(ABLE) IF a relation Rxy [e.g. x looks at y] in the SD
CORRESPONDS TO
a relation R*xy [e.g. x thinks about y] in the TD
THEN the relation x is-able-to-stand-in-R-to y [e.g. x is able to look at y] in the SD
CORRESPONDS TO
the relation x is-able-to-stand-in-R*-to y [e.g. x is able to think about y] in the TD.

Further adjuncts handle equally generic enabling, facilitating, and causal relations (enabling or causing x to V-y, or facilitating this activity or achievement). E.g.:

(CAUSE) IF a relation Rxy [e.g. x looks at y] in the SD
CORRESPONDS TO
a relation R*xy [e.g. x thinks about y] in the TD
THEN the relation z causes-Rxy [e.g. z causes x to look at y] in the SD
CORRESPONDS TO
the relation z causes-R*xy [e.g. z causes x to think about y] in the TD.[4]

Further generic adjuncts generate correspondences for inferences about the manner in, and extent to, which something is done or achieved (easily/with difficulty, intentionally/ accidentally, wholly/partly, well/badly, etc.), as well as about temporal attributes (order and duration of events, intermittence or persistence, rates of change, etc.) and the emotional and other valence attaching to the property or relation mapped:[5]

IF a relation Rxy in the SD [e.g. x looks at y]
CORRESPONDS TO
a relation R*xy in the TD [e.g. x thinks about y]

(MAN) THEN for any manner M: M(Rxy) [e.g. x carefully looks at y]
CORRESPONDS TO: M(R*a,y) [e.g. x carefully thinks about y]

(EXT) THEN for any extent E: E(Rxy) [e.g. x sees enough of y]
CORRESPONDS TO: E(R*xy) [e.g. x knows enough about y]

(T-ATT) THEN for any temporal attribute TA: TA(Rxy) [e.g. x persistently looks at y]
CORRESPONDS TO: TA(R*xy).[6] [e.g. x persistently thinks about y]

(VAL) THEN for any moral, emotional, or other valence V: V(Rxy) [e.g. x angrily looks (‘glares’) at y]
CORRESPONDS TO: V(R*xy) [e.g. x angrily thinks about y]

---

[4] Throughout, variables x, y, z… do not range only over individuals. They can take any fillers of the subject- and patient-roles of the relevant verbs as values.

[5] (MAN) and (EXT) below simplify formulations in the extant literature.

[6] Where TA actually amounts to a temporal relation (‘before’, ‘after’, ‘until’, etc.), a further correspondence is required as input: (T-REL) If $R_1xy \leftrightarrow R_1$*xy and $R_2xy \leftrightarrow R_2$*xy, and R_1xy stands in temporal relation T to R_2xy, then R_1*xy stands in T to R_2*xy.

Wherever our SD premises attribute a valence, temporal attribute, extent, or manner to something that gets mapped into the TD, analogical inferences with these correspondences project them too into the TD. All such projections are defeasible.

Conceptual metaphors that could be obtained through such a minimal-analogy strategy, namely by building up from a given core mapping with a restricted range of generic mapping adjuncts, are what we called *'narrow conceptual metaphors'*. Thus the set of correspondences we can build up to from core correspondences (1) and (2), respectively, are constitutive of the narrow conceptual metaphors KNOWING AS SEEING and THINKING-ABOUT AS LOOKING-AT, respectively.

Where initial source-domain reasoning yields conclusions that employ both concepts which are mappable with narrow conceptual metaphors and generic concepts that apply in both SD and TD, the conceptual metaphors can be complemented by mappings of these generic elements. Since they obtain in both domains, they get mapped onto themselves as a default, in mapping governed by semantic similarity (see above; cp. Forbus et al. 1995, 2016). Narrow conceptual metaphors thus get complemented by *generic self-mappings* like, e.g.:

(U) S uses X ↔ S uses X

Generic mapping adjuncts can then also be applied to these correspondences.

Next, we outline how these restricted mappings can be deployed to derive interpretations for metaphorical expressions that form part of extended metaphors (Section 3). Then we will consider how the computationally implemented strategy can contribute to an empirical account of how (some) metaphors are understood (Section 4).

3. Interpreting Metaphors: Applying ATT-Meta

The ATT-Meta approach uses a three-step procedure for interpreting metaphorical expressions in sentences: First, it interprets the expressions literally and makes from the literally interpreted sentence inferences that deploy general knowledge about the source domain. This *source-domain reasoning* may involve abstract re-representation of the initial premises. It delivers conclusions that are mappable from SD to TD with the modest resources we have just reviewed. In a second step, the expression at issue is treated as metaphorical. In line with fictionalist accounts of metaphor (e.g., Walton 2004), this is cashed in as treating the sentence and the conclusions derived from it in the first step as a piece of fiction and developing a 'pretence scenario' in which we 'pretend', e.g., that a thinker is literally looking at an option or issue (in something like the way in which fairy tales pretend that pots talk to kettles): Precisely to prevent nonsensical conclusions, these sentences are placed in a 'pretence cocoon' from which only restricted analogical inferences about 'reality' or the intended target-domain application are allowed. These *restricted analogical inferences* eschew generation, involve only substitution, and make do with the restricted range of mappings we have just reviewed. Third, subject to contextual constraints, one or more conclusions of such inferences are then chosen as interpretation that specifies the utterance content. Especially where prior abstract re-representation was involved, the conclusion of the analogical inference may first be rendered more specific through *target-domain reasoning*. The initial source- and final target-domain

reasoning deploys only knowledge or assumptions that are generally shared, and generally taken to be so shared.

Let's consider how this approach can be applied to deliver interpretations for vision-cognition metaphors that are generally acknowledged as philosophically highly important but have received only rather little and mostly cursory discussion in the extant conceptual metaphor literature (Danesi 1990, Goschler 2005, Kövecses 2010, Lakoff & Johnson 1999, Sweetser 1991). They still lack a detailed analogical analysis (with ATT-Meta or any other approach).

For an initial understanding of the subtle ATT-Meta approach, consider how it can be used to derive metaphorical interpretations for the vision term 'clear'.[7] First, it interprets this expression literally, and makes elementary stereotypical inferences: As the *Macmillan English Dictionary for Advanced Learners* (*MEDAL*) explains, 'clear' literally means 'easy to see' (sc.: for somebody or other). An elementary *source-domain inference* has it that

(SI) When X is literally clear (i.e. easy to see), then

(SC) any [contextually relevant] subject S easily can see X.

An *analogical inference* then takes us from this source-domain conclusion (SC) to a target-domain conclusion

(TC) Any [contextually relevant] subject S easily can get to [*sic*] know X

To obtain the correspondence for this inference, we first apply the ability-adjunct (ABLE) to core mapping (1) and then the manner-adjunct (MAN) to the resulting correspondence. This illustrates how the use of specific adjuncts can subtly influence the meaning of the mapped core expression: By 'knowledge' we ordinarily understand a comparatively stable or persistent state that may result from an intellectual effort or achievement. (ABLE) highlights the achievement aspect of 'seeing': S is able, manages to see X. It hence has us map 'can see' on the ability to bring off an epistemic achievement: not on 'can know' but on 'can *get to* know'. (MAN) then transfers 'easily', and we obtain:

Any [contextually relevant] subject S easily can see X

↔ Any [contextually relevant] subject S easily can get to know X

This correspondence is used for the analogical inference from the conclusion (SC). Together, the analogical and prior source-domain inference make up a simple inference chain. The final conclusion of this chained inference, viz. (TC), then specifies a metaphorical interpretation of the expression from which we proceeded, namely, in the initial premise of the source domain inference SI ('X is clear'). We thus obtain the default metaphorical interpretation for 'X is clear' that is reflected in the dictionary explanation 'manifest to judgment, plain, evident' (*Oxford English Dictionary*, *OED*).

In shifting the correspondence to an epistemic achievement, (ABLE) allows for different closely related achievements, including 'can understand'. Applying (MAN) to this correspondence, instead, has us move from (SC) to

[7] Here and throughout this paper, we deploy the ATT-Meta theory and forward-reasoning to obtain default interpretations for sub-sentential expressions or open sentences. The computational implementation of the theory (also called 'ATT-Meta') actually employs goal-directed reasoning and interprets whole sentences.

(TC') Any [contextually relevant] subject S easily can understand X.

This conclusion captures the closely related interpretation informing another dictionary explanation: 'easy to understand' (*MEDAL*).

To forcefully bring out how the ATT-Meta approach economises on mappings, while delivering rich interpretations, consider how we can use it to deliver metaphorical interpretations for the expressions 'beyond my ken' and 'focus'. The *OED* explains the (now rare) literal sense of 'ken' as 'range of sight or vision'. Standard conceptual metaphor theory would then posit a mapping from 'ken (range of vision)' to 'range of knowledge or understanding'. ATT-Meta, by contrast, proceeds from a source domain inference: When something is beyond someone's range of vision, he is typically unable to see it. I.e.:

(SI) If X is literally beyond the ken of S, then

(SC) S is unable to see X.

An analogical inference then takes us from (SC) to the target-domain conclusion

(TC) S is unable to understand X.

This analogical inference does not require correlating a further element of the visual SD ('ken') with a TD element. Rather, the necessary mapping can be derived from the core mapping (1) of 'seeing' onto 'knowing', by applying the ability-adjunct (ABLE). This stresses the achievement aspect of 'see' and has us correlate the ability to see with the ability to pull off the achievement of getting to know or understand (cp. above). Applying (NEG) to the result gives us the correspondence:

S is unable to see X ↔ S is unable to understand X

The target domain conclusion (TC) thus obtained provides a default metaphorical interpretation of 'X is beyond the ken of S'.

A richer interpretation can be obtained by taking into account that the present inability to see has a particular cause: It is not due to blindness or darkness. Rather,

(SI) If X is beyond the ken of S,

(SC) S is unable to see X because S does not see far enough.

To map this richer conclusion, ATT-Meta needs to re-represent it in more abstract terms:

(SC') S is unable to see X because S does not see to a sufficient extent.

By applying (EXT) to core correspondence (1), we obtain 'S sees to a sufficient extent ↔ S knows to a sufficient extent (has enough knowledge). (Since (EXT) does not stress the achievement aspect of 'seeing', the correlation is with a state of knowledge, rather than an epistemic achievement.) We then apply (NEG) to the result, and finally (CAUSE) to the present and previous input, and thus obtain a correspondence that underpins the analogical inference to

(TC') S is unable to understand X because S does not know enough.

This conclusion can be rendered more specific by invoking the target-domain knowledge that the presently relevant knowledge may be propositional or experiential. The resulting richer interpretation is articulated by this dictionary entry: 'impossible for someone to understand because they don't have enough knowledge or experience' (*MEDAL*).

In the ATT-Meta model, derivations of metaphorical interpretations may proceed from more than one conclusion of source-domain reasoning and can involve different core mappings, as in our next case: 'to focus'. With (1) 'see' and (2) 'look at', this probably is one of the three vision verbs most commonly used metaphorically.[8] Conceptual Metaphor Theory would invoke a further correspondence akin to (1) see ↔ know, and (2) look at ↔ think about, e.g. 'visually focus' ↔ 'mentally focus'. With the ATT-Meta strategy, we instead spell out a consequence of the literal interpretation of 'S focuses on X', which *MEDAL* explains with the words 'if you focus your eyes, you look at something carefully until you can start to see it clearly.' This articulates an elementary source-domain inference which seizes on semantic features of the verb:

(SI) If S focuses [her eyes] on X,
  (i) S looks at X carefully
  until
  (ii) S sees X well.

By applying the adjunct (MAN) to core mapping (2), we obtain a correspondence between consequent (i) and 'S thinks about X carefully'. Applying the same adjunct to core mapping (1), yields a correspondence between (ii) and 'S knows well what X is'. Indeed, since the evaluative term 'well' highlights the achievement aspect, it invites a correspondence with 'S understand X well'. The two correspondences for (1) and (2) and the temporal relation 'until' provide input for (T-REL) (Fn.6) which takes us from the source-domain conclusion to 'S thinks carefully about X until S knows well what X is' – or understands X properly. In its third and final step, the strategy has us rely on target-domain knowledge to flesh out the above interpretation of 'S focuses on X', namely, by spelling out various ways in which one may 'think about' something in soliloquy, debate or writing, to obtain the interpretation: 'To carefully reason about or discuss X, until one understands X properly'. To interpret metaphorical talk of 'focussing on' something, our approach hence does not add another correspondence to those for 'see' and 'look at', but derives new more specific correspondences from those core correspondences, with a couple of generic adjuncts that belong to a limited range of such adjuncts.

4. Towards a Minimal-Analogy Theory

So far we have described a computationally implemented strategy for deriving metaphorical interpretations, and demonstrated how it can be applied to derive interpretations for extended metaphors. We will now consider how this ATT-Meta model coheres with psycholinguistic accounts of language comprehension and how it can contribute to a theory of how extended metaphors are actually processed and understood, in ordinary discourse. We will thus build up towards a theory that explains how such metaphorical expressions are initially processed and understood in actual discourse. This new *'Minimal Analogy Theory' (MAT)* of extended metaphor explains how rich interpretations of such metaphors are obtained through minimal use of restricted analogical resources, and their interaction with routine comprehension

[8] These three verbs jointly account for 85% of metaphorical uses of sight terms in a corpus obtained from a naturalistic context (verbal lecturer-student interactions) (McArthur et al. 2015).

processes that are empirically well attested: According to MAT, initial interpretation of such metaphors involves routine stereotypical and predictive inferences, followed by restricted analogical inferences (Section 3) and, where necessary, by integration with antecedent world knowledge (about the TD) and standard pragmatic inference (see Section 5).

In psychological reality, the initial source-domain reasoning envisaged by ATT-Meta typically involves the sort of largely automatic inference processes that are supported by associative processing in semantic memory (McRae & Jones 2013, Neely 1991) and routinely go on in language comprehension: semantic and stereotypical inferences triggered by individual words or phrases (*'stimulus-driven inferences'*) (Hare et al. 2009, Harmon-Vukic et al. 2009) and 'expectation-driven' *predictive inferences* from prior text and world-knowledge (Metusalem et al. 2012, Ratcliff and McKoon 1989). Conclusions or outputs of these initially parallel processes get subsequently integrated (Giora 2003, Peleg et al. 2004, Peleg & Giora 2011): Where they can contribute to the interpretation, they are retained (Giora & Fein 1999, Fein et al. 2015); where they interfere, they are effortfully suppressed (Gernsbacher & Faust 1991, Faust & Gernsbacher 1995) (*'Retention/Suppression Hypothesis'*, Giora et al. 2014). Retained conclusions can serve as premises for subsequent analogical inferences.

Many nouns (Hare et al. 2009, McRae et al. 2005) and verbs (Harmon-Vukic et al. 2009, Ferretti et al 2001, McRae et al. 1997) are associated with *stereotypes*: sets of features that come to mind first, and are easiest to process, when we hear those expressions. In psycholinguistics, such associations are often identified through sentence-completion, listing, and plausibility ranking tasks (McRae et al. 1997). Their strength is measured through the 'cloze probability' or frequency with which a feature is named in a sentence-completion task like 'Elephants are___'. Nouns are stereotypically associated with the most frequently observed or talked-about properties of their bearers (elephants are clumsy and large, and have phenomenal memory). Verbs can be associated with more complex, internally structured stereotypes, aka *'generalised situation schemas'* (Rumelhart 1980), made up of typical features of the relevant events or actions, agents, and patients (i.e. referents of direct objects). (E.g.: 'She manipulated Joe. He is so____'– gullible, naïve, stupid. 'Jack was manipulated by Jane. She is so___' – cunning, shrewd, clever.) When competent language users encounter these expressions in sentences, they automatically infer stereotypically associated attributes and consequences, in line with the neo-Gricean I-heuristic: 'Find interpretations that are stereotypical and specific!' (Levinson 2000). Together with semantic inferences, these massively parallel inferences constitute the bulk of source domain reasoning which may preface analogical inferences, in metaphor interpretation.

In fact, stereotypical inferences facilitate metaphor interpretation both with and without analogical reasoning. They also facilitate attributional metaphor interpretation strategies (Bortfeld and McGlone 2001, Searle 1993), which require no analogical reasoning: When hearing 'Achilles is a lion', you will automatically infer that Achilles is strong, ferocious, brave, and noble. In some contexts ('The zoo calls its giraffe 'Hugo' and…'), these stereotypical conclusions will be used to *enrich* literal interpretations through pragmatic inferences that can be immediately cancelled ('but the poor animal has grown weak and miserable in captivity'). In other contexts ('According to the *Iliad*…'), one or more of these conclusions will be taken to *constitute* the interpretation or intended meaning. The contextually

inappropriate literal attribution (of lionhood) is suppressed and replaced by that of one or more stereotypically associated properties selected as interpretation (e.g., strength and nobility). The property selection process involved builds on pertinent background knowledge (the *Iliad's* Achilles is a human warrior) but is highly sensitive to context (as in this example from Wallington 2010):

(a) Mary is graceful, but John is an elephant.
(b) Patricia is small, but James is an elephant.
(c) Susan forgets everything, but Paul is an elephant.

The multiplicity of stereotypically associated properties can account for the indeterminacy and context-sensitivity of the metaphorical use of the word. Arguably, in an appropriate context, any property can be selected in this way, if sufficiently strongly stereotypically associated with the metaphorically used word.[9]

Also predictive inferences can support metaphor interpretation with and without analogical inference: When reading that an elephant or a bull is in a China shop, readers will not only infer that the animal is clumsy and bulky, and that the place is full of fragile objects, but also predict that the animal is liable to break many fragile things. We exploit this inference for metaphorical extension when we talk of someone being 'a bull' (in English) or 'elephant' (in French, German, or Italian) 'in a China shop'. Depending upon context, the inferred attribute (X is liable to break many fragile things) may be applied literally ('Amidst the delicate furniture…') (*attributional metaphor*). In other contexts ('During the difficult negotiation…'), the inferred source-domain conclusion merely provides the basis for analogical inference that delivers the intended interpretation (*analogical metaphor*) (Bortfeld & McGlone 2001).

The Minimal Analogy Theory (MAT) accordingly takes the kind of analogical inferences specified by ATT-Meta to be involved only in the latter case[10] – and only where the metaphorical uses at issue are comparatively new to the hearer. According to the influential *Career of Metaphor Hypothesis* (Bowdle & Gentner 2005), metaphorical uses of expressions prompt a distinctive interpretation process only when they are new to language users, and get processed like literal uses, once they have become familiar and contributed to building up a new category. According to the empirically well-supported *Graded Salience Hypothesis* (Fein et al. 2015, Giora 2003), semantic and stereotypical features associated with frequent and familiar uses of an expression get initially activated upon encounter of the expression, regardless of context – and of whether the use at issue is literal or figurative. These features

[9] The likelihood of selection is, however, not merely a function of strength of stereotypical association and degree of contextual fit: Whereas the properties appearing in generic mapping adjuncts are selected as a default, others, like colour, need not be selected even when contextually appropriate ('Mary's pencil is blue but John's is a tomato', Wallington 2010) and seem to be selected only together with other properties: E.g., 'emeralds of your face' and 'pearls of your mouth' (Herrero 2003) readily conveys information about both colour and (aesthetic) value (beautiful green eyes or white teeth). Discussion of contextual and other constraints on property selection is, however, beyond this paper's remit.

[10] By contrast, the ATT-Meta model, which has been developed not for psychological explanation but to deliver interpretations for as many metaphorical uses as possible, delivers interpretations for both attributional and analogical metaphors, and employs analogical inferences also for the former purpose. See, e.g., Wallington (2010).

jointly form generalised situation schemas (Rumelhart 1980). Repeated analogical inferences can build up such a schema which will subsequently be directly activated by the verbal stimulus (without 'analogical detour') and deployed for categorisation judgments (Tversky & Kahneman 1983) in the same way as other schemas (Bowdle & Gentner 2005). Accordingly, initial comprehension inferences will be followed by analogical inferences only when language users encounter metaphorical uses of expressions that have not yet become familiar to them.

ATT-Meta specifies, and MAT invokes, an analogy strategy which builds up from a few familiar core mappings, with widely applicable generic mapping adjuncts. Inferences in line with this strategy are particularly well suited to explain the wholesale metaphorical extension of entire families of related expressions from more concrete to more abstract domains (e.g. the systematic recruitment of visual terms for talk about intellectual activities and achievements) and their ready extension through apparently unrelated terms (e.g., 'bury' for visual metaphors: When S buries X, she makes it impossible for people to see/know X, and prevents people from looking at/thinking about X). We will use MAT to explain the interpretation of extended metaphors.

The extended metaphors we are interested in are all linguistically realised through metaphorical uses of entire families of related expressions, which have become conventionalised to the point of finding entry into dictionaries. In contrast with stereotype-based attributional metaphors which often are highly sensitive to context (above), these expressions have *default metaphorical interpretations:* interpretations which language users predictably give expressions, as and when they initially encounter their metaphorical use; these default interpretations are modified or dropped only in the light of further contextual information or social feedback.[11] In the absence of such modification, repeated analogical inference will build up a new category (Bowdle & Gentner 2005) or, more specifically, a new generalised situation schema (Rumelhart 1980), which will subsequently be directly activated by the verbal stimulus. This schema need not be associated with another expression, or may be associated with its use in only one of several senses.[12] In either case, the specification of the metaphorical interpretation will require more than a one-word paraphrase. These potentially rich interpretations will be implicitly presupposed in fast-paced conversation. In unhurried contexts of 'metaphor appreciation' (Gerrig & Healy 1983) where competent language users judge the aptness of metaphorical expressions, these interpretations are not only made explicit but can also be developed further, in predictable ways. MAT seeks, first, to specify the potentially rich and complex default interpretations presupposed in ordinary discourse, second, to explain how they are initially obtained and, third, to predict how they will be developed further.

We will now focus on the first task. Default metaphorical meanings stand a better chance of widespread conventionalisation than interpretations that require specific and historically contingent real-world knowledge (cp. Traugott & Dasher 2005). They are made explicit, e.g.,

---

[11] For a review of related but distinct notions of 'default interpretation', see Jaszczolt (2011). Cp. Giora et al. (2015).

[12] E.g., the schema associated with "keep in mind" (see Section 5) corresponds to one sense of "remember" (MEDAL sense 2a), but not others (e.g., MEDAL sense 1).

when language teachers and students explain their understanding of metaphorical expressions (Bortfeld 1998) – and by 'advanced' dictionaries. If we assume that most of the expressions belonging to extended vision-cognition and space-cognition metaphors have kept their default meanings through conventionalisation, we should therefore expect their dictionary explanations to reflect rich default interpretations that cannot be captured by a single concept but can be derived by MAT. I therefore propose to test this variant of ATT-Meta by verifying that it can generate the interpretations given in the *Oxford English Dictionary* (*OED*) or *Macmillan English Dictionary for Advanced Learners* (*MEDAL*), whichever gives a richer explanation.[13]

5. Metaphorical Minds

Talk of 'minds' in ordinary discourse revolves around (though it is not limited to) spatial cognition metaphors. We will now verify that the Minimal Analogy Theory (MAT) outlined delivers accurate interpretations for such ordinary mind-talk, when working in conjunction with well-attested language processes like pragmatic strengthening (Levinson 1983, Traugott 1989). This will, first, support the hypothesis that what analogical reasoning is employed in interpreting this pre-philosophical talk uses an analogy-minimising mapping strategy. Second, it will reveal a surprising fact about the place of 'minds' in the analogical reasoning that underpins pre-philosophical metaphorical talk. This finding will provide the basis for exposing (in section 7) a specific fallacy in philosophical reasoning about the mind (reconstructed in section 6).

Conceptual metaphor theorists quite unanimously regard the use of English 'mind'-idioms as motivated by a conceptual metaphor that treat minds as target-domain entities and correlates them with containers in the source domain (MIND AS CONTAINER) (e.g., Gibbs & O'Brien 1990, Koivisto-Alanko & Tissari 2006, Kövecses 2010, Lakoff & Johnson 1980, 1999). We will now explore a rather different new approach which maps a spatial relation, rather than a spatial entity (like a container), and maps it onto a cognitive relation, rather than any target-domain entity ('mind', or some such).

Much metaphorical 'mind'-talk is grounded in the pretence or fiction that every thinker has a personal physical space or container. But this space or container does not get placed into correspondence with anything we could conceptualise as an element of the intellectual TD – say, with our 'rational or intellectual powers' (as the *OED* explains another use of 'the mind'). Indeed, in metaphorical 'mind'-talk about what people think of or remember, the fictitious space or container does not get placed into correspondence with *anything*. In the expressions of interest, 'the mind' rather serves as label for the fictitious space (rather than any TD correlate) a *relation to which* gets mapped onto a cognitive relation, by *core mapping(I):*

X is *inside* a physical space belonging to S ('inside the mind of S') ↔ S thinks of X.

ATT-Meta theorists have stressed that the SD reasoning involved in metaphor interpretation often involves elements that do not get mapped onto the TD (e.g., Barnden 2015). The physical space figuring in (I) is a case in point: Neither (I) nor any correspondence MAT generates from

[13] For discussion of various problems involved in this use of dictionaries, see Steen et al. (2010).

it places this space (‘mind’) into correspondence with anything; rather, the spatial relation ‘X is in the mind-space of S’ gets correlated with something, namely, with the cognitive relation ‘S thinks of X’. Generic mapping adjuncts generate further mappings from this core mapping (I). Together, these mapping are constitutive of the narrow conceptual metaphor BEING THOUGHT OF AS BEING IN A PERSONAL SPACE.[14]

We will now see how this restricted mapping allows us to derive interpretations for ordinary talk that combines ‘to/from/in/ the mind’ with verbs including ‘spring’, ‘come’, ‘cross’, ‘bring’, ‘call’, ‘bear’, ‘keep’, have’, ‘put in/out’, and ‘banish’. For all these complex expressions, we can derive default metaphorical interpretations in line with MAT: by prefacing such restricted analogical inferences with the most elementary source-domain inferences (SI) which merely make explicit semantic or stereotypical implications of the verbs employed alongside ‘the mind’, and – sometimes – developing initial target-domain conclusion further, with standard pragmatic inferences.

As first example, consider ‘X springs to S’s mind’. To derive its default interpretation, we first interpret the phrase as being literally about entry into a physical space, and bring out some implications. The verb ‘spring’ implies a certain suddenness and that the outcome results from action of the subject-role filler, rather than the patient-role filler (here: X, not S):

(SI) When X springs into S’s space (mind),
  i. X suddenly is in the space of S, without an effort on the part of S and previously
  ii. X was not in the space of S

To obtain a mapping for analogical inference from (i), we start with mapping (I) and apply to it the mapping adjuncts (T-ATT) and (MAN), which carry over ‘suddenly’ and ‘without effort’, respectively. This secures correspondence of (i) with ‘S thinks of X suddenly and effortlessly.’ For inference from (ii), we apply the mapping adjuncts (NEG) and (T-ATT) to (I) and thus get the correspondence between (ii) and ‘S did not think of X’.[15] These two correspondences for (i) and (ii) and the temporal order-relation ‘previously’ provide input for (T-REL) (Fn.6), that correlates the entire consequent of (SI) with ‘S suddenly and effortlessly thinks of X, and previously did not think of X.’ As our dictionaries put it, ‘to spring to mind’ is ‘to occur immediately to a person, be one’s first or instinctive thought’ (*OED*); ‘you suddenly start to think about it’ (*MEDAL*).

The core correspondence (I) is no more precise than our use of the verb “to think of”: We use it not only to speak of occurrent thought but also when we think of somebody or something continually, rather than continuously, or even just very occasionally.[16] It often gets

[14] This metaphor differs also from the relation-to-relation mapping the extant ATT-Meta literature proposes to account for similar linguistic data (cp. Fn.3). Comparative evaluation of these two approaches has to be reserved for another occasion.

[15] This involves a promissory note insofar as the ATT-Meta literature has not yet provided resources to handle tense. We simplify by taking past and future tense as temporal attributes projected by (T-ATT).

[16] Even in the most sincere love letter, 'I am thinking of you night and day' means that the writer thinks of the addressee again and again, rather than without interruption. And I may be ‘thinking of

disambiguated by implications of the verbs used in the metaphorical expressions at issue: E.g., ‘spring’ implies such suddenness that we can only delineate the point of change with sufficient precision in case the thought at issue is occurrent.

The expression ‘come to mind’ is used almost interchangeably (so that *MEDAL* gives the same explanation for both). However’ ‘come’ lacks the implication of suddenness and does not make the effortlessness on the part of S so salient. Hence:

(SI) When X comes into the space (mind) of S,
  i. X is in the space of S
  and previously
  ii. X was not in the space of S

Analogical inference from (i) requires only mapping (I), while inference from (ii) is as before, to yield with (T-REL) the interpretation: ‘S thinks of X after not thinking of X previously’, i.e., ‘S starts to think of X’. Again, the point of change can only be delineated in case the thought at issue is occurrent.[17]

The *OED*, however, offers a richer interpretation: ‘to occur (esp. upon reflection)’. We can explain the enrichment (‘upon reflection’) by pragmatic considerations (Levinson 1983): ‘come’ does not imply suddenness, and makes the agency of the subject-role filler less salient. Since the more informative expression ‘spring to mind’ is available, and the less informative ‘come to mind’ is no briefer, we infer from preference of ‘come’ over ‘spring’ that the implications not shared (or not shared to the same extent) by ‘come’ are not meant to apply: S starts to think of X, but does so neither suddenly nor effortlessly. I.e.: S starts to think of X upon effortful reflection, though the immediate trigger is still no action of S (as in a conscious logical derivation of X): ‘X occurs to S upon reflection’.

Straightforward reasoning applies to ‘X crosses the mind of S’: When X crosses a physical space, it is currently in it, typically was previously outside it, and will again leave it. Typically, a space ‘crossed’ is small by comparison to the entire trajectory of X. Therefore X will be in the personal space of S only for a comparatively short time. Mapping (I) and (T-ATT) facilitate inference from the conclusion of source-domain reasoning ‘X briefly is in the space belonging to S’ to ‘S briefly thinks of X’ (as the *OED* puts it: ‘(of a thought) occurring to one, esp. transiently’). Again, the temporal attribute enforces an occurrent reading of ‘think of’.

When someone ‘brings’ something to a location, he causes it to be there (which may but need not have been his aim), and there is the implication that it was not there before:

(SI) When X brings Y to the space (mind) of S
  i. X causes Y to be in the space of S
  and before then
  ii. Y was not in the space of S

---

travelling to Japan’ for months, even when I hardly ever find the time to devote thought to travel plans.

[17] Very similar reasoning lets us derive the conventional metaphorical interpretation for the now obsolete ‘to pass from / out of mind’: to be no longer thought of / remembered, to be forgotten (cp. *OED*).

Applying the (CAUSE) adjunct to Mapping (I) provides the mapping for analogical inference from (i) to 'X causes S to think of Y'. Derivation of 'and S did not think of Y before' then follows as above, to yield: 'X causes S to think of something Y which he did not think of before' (*OED:* 'to cause one to remember someone or something') as interpretation of "X brings Y to S's mind". We can also cause a person to be at a location she wasn't at before, by calling her to it: Parallel reasoning leads to the same interpretation for 'X calls Y to S's mind'.

Similar reasoning lets us interpret talk of 'putting things out of one's mind':

(SI) When S puts X out of the space (mind) of S
  i. S deliberately causes X not to be in the space of S
     and before then
  ii. X was in the space of S

To obtain the mapping required for inference from (i), we apply first (NEG) to mapping (I) and then (CAUSE) to the resulting mapping, while (MAN) carries over 'deliberately', so that we obtain: 'S deliberately causes S not to think of X.' Further derivation from (ii) and the temporal relation is by now obvious. The TD conclusion ('S has been thinking of X but deliberately caused herself not to think of X') can be paraphrased succinctly: S deliberately 'forgets about somebody or something, even if only for a short time' (MEDAL).

Talk of 'banishment' (as in 'The news of her pregnancy banished all other thoughts from her mind') has even richer implications:

(SI) When somebody or something X banishes Y from the space (mind) of S
  i. Y previously was in the space of S
  ii. X deliberately causes Y not to be in the space of S
  iii. X will prevent Y from being in S's space again (in the foreseeable future).

A mapping for inference from the genuinely new element (iii) is obtained by applying a prevent adjunct that works like (CAUSE) to core mapping (I), and (T-ATT) to the result, to obtain: 'X will prevent S from thinking of Y again (in the foreseeable future)' – and not only for a short time.

The interpretation of 'keep' and 'bear in mind' then illustrates the combined use of core mappings and generic self-mappings (Section 2): According to the *OED*, to 'keep' literally means to 'store in a regular place', namely, 'for future use'. ('To store' is explained as 'keep for future use'.) That S keeps X for future use (defeasibly) implies that S can make use of X and will make use of X, as and when required. (Why else bother to store it?)

(SI) When S keeps X in the space (mind) belonging to S,
  i. X is in the space of S
  ii. S can use X
  iii. S will use X, as and when required.

Analogical inference with mapping I from (i) yields 'S thinks of X'. (ii) and (iii) employ the generic use-relation, in which users can stand not only to physical goods but to any (other) object of thought as well. These denizens of both SD and TD are mapped by the generic self-mapping (U). Application of (ABLE) to it provides a self-mapping for (ii). Two-fold application to (U) of (T-ATT), for future tense and temporal qualifier, yields a self-mapping for (iii). Together, these mappings license analogical inferences that take us out of the pretence

cocoon and to conclusions about the TD: ‘S thinks of X, can use X, and will use X, as and when required.’ This conclusions is rendered more precise by taking into account contextual information, in pragmatic inference. (ii) and (iii) are incomplete: used for what? As and when required for what? Relevant contextual information is provided by (i): We are talking about thinking. The use at issue is hence the use in thinking, i.e., taking X into account. Pragmatic inference in line with Grice’s maxim of quantity takes us from (ii) and (iii) to the conclusion that S is not currently using X in his thinking and, therefore, the interpretation of ‘thinks of’ as ‘thinks of every now and then’ (rather than ‘has the occurrent thought’). The resulting interpretation is consistent with the dictionary explanations ‘to remember, not forget, take into account’[18] and ‘to remember something, especially something that will be important in the future’ (*MEDAL*).

The most salient sense in which you can literally ‘bear’, i.e., ‘carry’ something in a space belonging to you is to carry it in the enclosed space of a container. When you carry something around with you in a container, you can typically take it out and use it, as and when required. Indeed, if you bother to carry it around, you typically will use it, as and when required. Accordingly, parallel reasoning leads from ‘S bears X in mind’ to the same conclusions as above, which can then be deployed as explained.

By contrast, that you ‘have’ something in your personal space implies only that (i) it is in the place and (ii) you can make use of it; but the phrase lacks the salient implication (iii) of storage for future use which is carried by ‘keep’ and ‘bear’ and suggests there is no current use. As in the case of ‘come’ vs ‘spring to mind’ (above), pragmatic considerations can therefore enrich the interpretation of the less informative ‘have in mind’: Its preference over the otherwise more informative ‘keep’ and ‘bear in mind’ warrants the pragmatic inference that the speaker means to rule out the suggestion that there is no current use, and seeks to convey the opposite: She wants to convey that S is not only thinking of S but *currently* making use of it in her thinking. This is consistent with the dictionary explanations of ‘to have in mind’ as ‘to think of, contemplate’[19] and ‘to recall and take into consideration, keep one's attention fixed upon’ (*OED*).

Reconstructing the derivation of default metaphorical interpretations for these expressions puts us into a position to clarify the status of ‘the mind’ in the analogical reasoning reconstructed. In the conceptual metaphor literature (e.g., Gibbs & O’Brien 1990, Koivisto-Alanko & Tissari 2006, Kövecses 2010, Lakoff & Johnson 1980, 1999), it is unanimously assigned as an element to the target domain of spatial cognition metaphors. This is understandable: In ordinary discourse, we often use ‘the mind’ to refer to an element of the intellectual TD, namely, to the faculty of reasoning and understanding (*OED* sense 21: ‘a person's cognitive, rational, or intellectual powers; the intellect’) which one may possess to various degrees (‘have a fine mind’) or – sadly – ‘lose’. It is also used metonymically to refer to people who possess this faculty (‘two great minds were in attendance’) (ibid.). But in

[18] http://www.oxforddictionaries.com/definition/english-thesaurus/mind (last accessed 12/11/2015)

[19] http://www.oxforddictionaries.com/definition/english-thesaurus/mind. Cp. the explanation of ‘having someone/thing in mind’ as ‘to *be* think*ing* of someone or something’ (note the tense, my italics), http://www.oxforddictionaries.com/definition/english/have-someone/thing-in-mind (both accessed 12/11/2015)

interpreting the present metaphorical expressions in the way outlined, ‘the mind’ is used exclusively to refer to an element of the spatial source domain, namely, to the physical space assigned to the subject S, by the fiction or pretence-scenario *from* which we are making analogical inferences.

For consider: The derivation process outlined delivers metaphorical interpretations for the expressions that are used in the initial premises of source-domain reasoning (SI), e.g. ‘If S focuses on X…’ (Section 3). These expressions are taken literally in this reasoning. In the case of expressions that combine physical action verbs like ‘cross’ or ‘spring’, ‘bring’ or ‘put’, ‘bear’ or ‘keep’, with potentially spatial prepositions like ‘to’, ‘in’, of ‘from, and the noun ‘mind’, this initial literal interpretation of the verb enforces spatial interpretation of the preposition and has us take the noun to refer to some physical space, in initial source-domain inferences. I.e., in SD reasoning, ‘the mind’ stands for a physical space, an element of the spatial SD or pretence scenario. In subsequent analogical reasoning, this element does not get mapped: It is neither placed in correspondence with the reasoning faculty that ‘the mind’ refers to in the above-mentioned literal use, nor with any other element of the TD of cognition. Only a spatial relation to this unmapped SD element gets mapped (by I). If one wanted to assign ‘the mind’ *that is invoked by spatial cognition metaphors* to one of the domains used in the analogical reasoning involved in interpreting them, we would have to assign it – against the majority opinion – not to the target but the source domain.

To sum up, the noun ‘mind’ has an independent literal application in the target domain of cognition. But, in metaphorical talk, it is recruited to stand for the physical space the pretence-scenario of spatial cognition metaphors assigns to thinkers. As cross-linguistic comparison reveals, this is an illustration of a more general strategy: In German, three different cognition terms, viz. ‘Sinn’ (sense), ‘Gedächtnis’ and ‘Erinnerung’ (both: memory), along with ‘Kopf’ (head, typically regarded as bodily seat of our reasoning powers) are recruited to stand for that unmapped space in the pretence scenario (SD) of spatial cognition metaphors interpretable through exactly parallel derivations (Table 1).

((Table 1 around here. Now p. 36))

6. Introspective Minds

Introspective conceptions of the mind, as articulated in early modern philosophy and culturally influential to this day, are frequently regarded as intuitive and part of common sense. Various philosophers have suggested that the intuitions at the root of these conceptions result from spontaneous analogical inferences with linguistically realised conceptual metaphors, crucially including spatial and vision cognition metaphors (Lakoff & Johnson 1980, 1999, Rorty 1980, Fischer 2011, 2014, cp. Wittgenstein 1933/2005). Without addressing philosophical examples, recent experimental studies from cognitive and social psychology as well as communication science suggest spontaneous analogical inferences with metaphors occur in problem-solving (Thibodeau & Boroditsky 2011, 2013, 2015) under conditions characteristic of much philosophical thought: high level of abstraction (Keefer et al 2014), greater psychological

distance (Jia & Smith 2013), low confidence in own target-domain understanding (Landau et al 2014), and low level of target-domain knowledge (Vandeleene et al, in prep).

We distinguished wide from narrow conceptual metaphors (Section 2) and hypothesised that while analogical reasoning involved in metaphor interpretation, as a default, employs narrow conceptual metaphors (*pace* standard conceptual metaphor theory), thinkers typically employ wide conceptual metaphors or correlation-maximising mapping strategies when deploying analogies for problem solving ('differential processing hypothesis'). We provided initial support for the more novel first part of the hypothesis through analysis of metaphorical 'mind-talk' and will now apply the less controversial second part to the analysis of classical philosophical efforts to solve the problem, or answer the question, 'What happens when we think?' These analyses will jointly expose two fallacies in analogical reasoning with metaphors that are at the root of introspective conceptions of the mind.

According to our differential processing hypothesis, analogical reasoning in response to such a task employs the full Copying with Substitution and Generation (CWSG) procedure (Section 2) and fuller source-target mappings. Influential texts from early modern philosophy of mind couch discussion of the operations of thought in visual terms and explicitly compare 'the mind' to a 'closet' or enclosed space in which 'pictures' are viewed (e.g., Locke 1700/1975, II.xi.17, cp. II.iii.1); i.e. the mind is compared to a restricted visual field, and 'the understanding' to 'the eye' (op. cit. I.i.1), the organ of sight.[20] When deployed in analogical reasoning, these comparisons translate into the two mappings:

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

These mappings evidently cannot be obtained with our analogy-minimising mapping strategy, by applying generic mapping adjuncts to core mappings of vision cognition metaphors like (1) S sees X ↔ S knows what X is, or (2) S looks at X ↔ S thinks about X. They are, however, generated when the analogy-maximising mapping strategy of structure mapping theory (SMT) is applied to generate vision cognition mappings from truisms about the visual source domain SD, given common-sense background knowledge about the intellectual target domain TD. SMT (cp. Section 2) 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 TD contains suitably related elements (Gentner and Markman 1997, 2005). This general mapping-rule leads to mapping N, in inferences from SD truisms such as:

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

The first verb in each sentence is mapped by core mappings (1) and (2) of different vision cognition metaphors (Section 2). The next verb, 'x uses y', stands for a generic relation that obtains in both the visual SD and the intellectual TD. In SMT, this relation is hence immediately mapped onto itself (Forbus et al. 1995). This leaves us looking for an element of

[20] See Fischer (2011, chs.1-3) for a fuller discussion of analogical reasoning in early modern texts by Boyle, Locke, and Berkeley. For philosophical context, see McDonald 2003.

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* or *seeing*-relations that get mapped onto *thinking-about* and *knowing*, respectively. The requirement of parallel connectivity hence has us look for something we use when we think or get to know things. Since we then use our wits, reason, intellect, or understanding – different labels for the same faculty (*OED*) – we thus obtain Mapping N: eyes ↔ understanding. Those who first think of 'the mind' as 'what we use when we think', will instead correlate 'the eyes' with 'the mind' – and move to N only in reasoning that correlates 'the mind' with something else, so that SMT's 1-to-1 mapping constraint obliges them to find another mate for our visual organ.[21]

The most salient alternative mapping is Mapping M. I submit this mapping is grounded in a conception of 'the mind' which we elaborate in reasoning with the conceptual metaphor BEING THOUGHT OF AS BEING IN A PERSONAL SPACE, namely, in the reasoning that motivates the prominent expressions 'to keep in mind' and 'to bear in mind': In interpreting these expressions, we conceptualise 'the mind' as a storage space of things we can make use of in thinking, the things we can remember, and thus know (Section 5). We then obtain mapping M when we align an explanation (or informative representation) of this concept with a basic explanation (or informative representation) of 'visual field':

(1) The visual field is the space in which the things are that the subject sees.
(2) The mind is the space in which the things are that the subject knows.

The SMT mapping strategy tells us to immediately correlate the concepts that apparently recur in both (1) and (2): the 'is' of identity, 'space', 'X is in Y', 'subject' etc. In reasoning with the conceptual metaphor KNOWING AS SEEING, we will also correlate 'seeing' and 'knowing', as per its core mapping. On this basis, mapping M, like N, is obtained by enforcing Parallel Connectivity. This requirement has us correlate the relata of the relevant relations: the objects of sight (the things the subject sees) with the objects of knowledge (the things the subject knows) – and the visual field (the space in which the objects of sight are located) with the mind (the space in which the objects of knowledge are located).[22]

Once the analogy-maximising strategy has put the new mappings M and N into place, the introspective conception of the mind is just a few analogical inferences away: Its intuitive key tenets can be obtained through 'full-blooded' analogical (CWSG) inferences with vision cognition metaphors when – and only when – the narrow conceptual metaphors KNOWING AS SEEING and THINKING-ABOUT AS LOOKING-AT (Section 2) are complemented with the mappings M and N which analogy-maximising mapping strategies deliver. Relevant CWSG inferences then proceed from source-domain truisms like 'When we look at things, things are before our eyes' (cp. Fischer 2014, 2015), as in Table 2:

((Table 2 around here. Now p. 36))

[21] This accounts, I submit, for the fact that many early modern texts use both 'the mind' and 'the understanding' to stand for both organ and field/space of inner perception – two different things (review: Fischer 2011: 35-40).

[22] Spoiler alert: Section 7 will expose fallacies in these mapping-operations.

From trivially true premises (P_1-P_4) we thus obtain substantive conclusions (C_1-C_4) (non-identical substitutions underlined, generated elements in italics):

P_1 When we look at things, things are before our eyes.

C_1 When we think about things, things *are before* our understanding.

P_2 When we look at things, things are in our visual field.

C_2 When we think about things, things *are in* our mind.

P_3 Things before our eyes are in our visual field.

C_3 Things *before* our understanding *are in* our mind.

P_4 When we look at things, we perceive things with our eyes, in our visual field.

C_4 When we think about things, we perceive[23] things with our understanding, *in* our mind.

These intuitions apparently generate the spatial relations 'X is before Y' and 'X is in Y' in the target domain and radically transform the notions of 'mind' and 'understanding'. When used on their own (rather than as part of complex expressions) in ordinary discourse, these words are primarily used to refer to intellectual powers or faculties (with further metonymical uses derivative from this primary use), namely, to 'a person's cognitive, rational, or intellectual powers [!]' (*OED*, sense 21 of 'mind') and her 'faculty [!] of comprehending and reasoning', aka 'intellect' (*OED*, sense 1 of 'the understanding'). The same holds true of philosophical discourse, where introspective conceptions replaced Scholastic, ultimately Aristotelian conceptions of 'souls' or 'psyches' as collections of powers and faculties (e.g. 'rational psyche' as powers of reasoning and volition, or 'sensitive psyche' as set of powers of perception, locomotion, and a-rational desire) (Bennett & Hacker 2003: 12-19). The present analogical reasoning reconceptualises what were previously sets of faculties (which cannot be meaningfully said to stand in any spatial relations) into a perceptual space and an organ of sense that peers into that space (both of which participate in spatial relations).

Crucially, *only* the new mappings N and M take us through vision cognition 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 (P_1-P_4) when we do not employ the new fare but make do with the narrow conceptual metaphors (1) KNOWING AS SEEING and (2) THINKING-ABOUT AS LOOKING-AT. We then get different conclusions which do not generate any spatial relations in the TD:

C_1* When we think about things, things are before our eyes.

C_2* When we think about things, things are within our visual field (ken).

C_3* Things before our eyes are in our visual field. (=P_3, for want of suitable mappings)

C_4* When we think about things, we perceive things with our eyes, in our visual field.

[23] Since 'perceive' (*OED*: '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 gets mapped onto, and substituted by, itself, in analogy-maximising reasoning with SMT-style mapping.

The remaining visual expressions "before our eyes" and "in our visual field (within our ken)" have default metaphorical interpretations with vision cognition metaphors. These do not even faintly suggest reference to any organ or space of perception. For "before our eyes" we can derive with MAT and KNOWING AS SEEING:

(SI) When something X is before S's eyes then

(SC) S can easily see X.

This stereotypical inference furnishes the premise for an analogical inference with a mapping we obtain from core mapping (1) with (ABLE) and (MAN) which leads to the conclusion that S can easily get to know or understand X (cp. interpretation of 'clear' in Section 3). The derivation for "within the visual field/ken" is even simpler:

(SI) When something X is within the ken of S,

(SC) S can see X.

This stereotypical inference furnishes the premise for an analogical inference – similar to that for 'X is beyond my ken' (Section 3) – that delivers the interpretation 'S can understand X'. We thus get these metaphorical interpretations:

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

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

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

'When we think about things, we get to know various things.'[24]

To sum up: analogical reasoning with vision cognition metaphors only gets us from source-domain 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 the further mappings M and N which are not part of those narrow conceptual metaphors. If we eschew these further mappings and apply our default analogy-minimising interpretation strategy, we obtain no conclusions that would even faintly suggest the conception of an inner organ and space of perception involved in thought.

7. Two Fallacies

We have reconstructed the analogical reasoning involved in interpreting ordinary metaphorical talk of 'minds' (Section 5) and in generating introspective philosophical conceptions of the mind (Section 6), respectively. Their comparative analysis allows us to expose seductive fallacies in the philosophical reasoning reconstructed. Analogical reasoning is governed by openly heuristic rules. Whereas *normative rules* determine or constrain what is correct, right or reasonable, *heuristics* are rules of thumb which yield reasonably accurate judgments in most relevant contexts, without constraining what is to count as correct. Such rules are never guaranteed to preserve truth. In talk about heuristic reasoning, the label 'fallacy' therefore tends to be reserved for cases where application of the relevant rules predictably leads from true premises or accurate information to conclusions or intuitive judgments that violate normative

[24] This interpretation involves SD inference from 'we perceive things with our eyes' to 'we see things', followed by analogical inference with core mapping seeing ↔ knowing. 'in the visual field' is ignored as redundant. (Where else would we see things?)

rules – think, e.g., of the 'conjunction fallacy' which arises from the use of the representativeness heuristic (Tversky & Kahneman 1983).

The normative rules most frequently referred to in the heuristics literature (reviews: Gigerenzer et al 2011, Kahneman 2011) are rules of logic and probability theory, whose violation results in judgments that cannot be true (together) or reasonable to accept. However, normative rules also include syntactic and semantic rules of language, whose violation results in 'nonsense' or conclusions which lack determinate meaning. I therefore propose to extend the notion of *'fallacy'* to cases where heuristic rules *predictably lead to conclusions that are semantically deficient* by lacking not (just) truth but determinate meaning. Indeed, as long as this deficiency can be predicted by study of the heuristic rules at issue, I want to say we are dealing with a 'fallacy', regardless of whether or not the deficiency is due to violations of normative rules. We will now identify two fallacies in this slightly more comprehensive sense. These fallacies, one general, the other specific, both frequently made in abstract reflection, occur in analogical reasoning with conceptual metaphors, namely, at the stage of mapping.

Let's first build up towards the general fallacy. It arises from the fact that, in analogical reasoning with conceptual metaphors, analogy-*maximising* mapping strategies (like SMT) used in problem-solving may have us make mappings that lead to conclusions which we cannot interpret with the analogy-*minimising* strategies we ordinarily employ in metaphor interpretation. Where the default reasoning strategy for analogical problem-solving employs conceptual metaphors and leads to conclusions we cannot interpret with the default strategy for metaphor interpretation, we are liable to be left with a claim whose meaning escapes us: Barring fortuitous semantic rescue, these conclusions lack determinate meaning.

The above conclusions C_1 to C_4 illustrate this point. In contrast with their starred counterparts, we cannot use the default interpretation strategy to derive metaphorical interpretations for them, with the vision cognition metaphors used to derive them: C_1 to C_4 employ at least one of two phrases we obtain when applying Mappings N and M to source-domain truisms: "before our understanding" and "in our mind". 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 vision cognition metaphors to map, and our default interpretation strategy of making restricted analogical inferences with narrow conceptual metaphors, from source-domain implications, gets no grip. Similarly, '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 analogical inference with a narrow vision cognition metaphor. The two key phrases lack default metaphorical interpretations with the vision cognition metaphors used to derive the relevant conclusions.

They also lack literal interpretations: In literal talk about the intellectual target domain, both 'the understanding' and 'the mind' ordinarily refer to faculties or powers of reasoning. Faculties and powers cannot be literally placed in spatial relations (like the generated relations 'x is before y' and 'x is in y'). Hence neither 'before our understanding' nor 'in our mind' can be interpreted literally, in target-domain talk. 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 default metaphorical interpretation with the conceptual metaphors used to derive them.

Other conceptual metaphors, or metonymies, may come to the semantic rescue: E.g., the core mapping (I) of the spatial memory metaphor (Section 5) lets us interpret the phrase 'in the mind', and this provides a readily intelligible interpretation for some conclusions (e.g., 'When we think about things, we think of things' for C_2), if not for others (e.g., 'Things before our understanding are thought of by us' for C_3). Similarly, spatial time metaphors (Gentner et al. 2002) may suggest to us a temporal interpretation for the phrase 'before the understanding', as 'prior to the act of understanding',[25] which may yield intelligible interpretations for some conclusions (though perhaps not C_1-C_4).[26] In the absence of such fortunate coincidences (and prior to ingeniously noticing and exploiting them), thinkers are unable to give determinate meaning and content to conclusions like C_1 to C_4. Early modern philosophical texts provide evidence for this inability in the shape of explanations of meaning which either remain purely negative or get disregarded almost the moment they have been given (see Fischer 2011, 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:

(C_1) *When* we *think about* X, it *is before* our understanding.
(C_2) When we think about X, it *is in* our mind.
(C_3) *When* X *is before* the understanding, it *is in* the mind.
(C_4) *When* an object of thought X *is perceived with* the understanding, it *is before* the understanding and *in* the mind.

Deeply integrated mappings endow analogical conclusions with high subjective plausibility (Gentner et al. 1993, Lassaline 1996). 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 simultaneous use of vision cognition metaphors and mappings M and N, which do not belong to the narrow conceptual metaphors employed in interpreting such metaphorical talk. These further mappings are pernicious insofar as they 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

[25] I am indebted to an anonymous referee for this suggestion.

[26] E.g., for C_1 a referee suggested the reading 'We think about things prior to understanding them'. However, while this claim is intelligible, its derivation requires interpreting 'understanding' as standing for a relation, viz. between subjects and objects of thought. Derivation from P_1 with mapping N treats the understanding as an entity (object or event), rather than a relation, even if a spatial-temporal mapping is applied to 'X is before Y'. A derivation with N would therefore not warrant the proposed reading.

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 that forms part of narrow vision cognition metaphors. M and N, however, 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 default metaphorical interpretation with vision cognition metaphors. They thus make E’s default interpretation with these conceptual metaphors impossible. In this sense, those mappings are *inconsistent with the default metaphorical interpretation of E with particular conceptual metaphors CM (‘default CM-interpretation’).*

Once metaphorical uses have become familiar or conventional, their interpretation no longer requires analogical inference (Bowdle & Gentner 2005). 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 may rather arise when our default strategy for analogical reasoning in problem solving is used in reasoning from SD premises which employ a complex expression E that has a default CM-interpretation: When we then make simultaneous use of the conceptual metaphor CM and mappings inconsistent with the default CM-interpretation of E, we will obtain a fresh conclusion that cannot be interpreted in line with our default interpretation strategy. 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 quite general and potentially hard-to-spot fallacy that may be committed at the mapping-stage of analogical reasoning. Let’s call it the ‘*metaphor-overextension fallacy*’. It consists in extending a narrow conceptual metaphor CM (such as, e.g., KNOWING AS SEEING) by adding mappings inconsistent with default CM-interpretations (like mappings M and N). The rules of ‘full-blooded’ 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 a narrow conceptual metaphor CM and mappings that are inconsistent with the CM-default interpretation of a complex expression employed in the premises.

The second fallacy exposed by our above reconstructions is more specific: It consists in a mis-mapping of the concept of ‘the mind’, in analogical reasoning from visual source domains. In its primary application in the intellectual target domain, ‘the mind’ stands for our power of thought (OED: ‘a person's cognitive, rational, or intellectual powers’), i.e., the reasoning faculty that allows us to get to know and understand things. In reasoning with the core correspondence (1) of ‘seeing’ with ‘knowing’, we ought to correlate this faculty, as a default, with the faculty that allows us to get to see things, viz. our sight (*OED* sense 8a: ‘the faculty or power of seeing, as naturally inherent in the eye’):

sight ↔ mind

Within our minimal analogy approach, we obtain this mapping by applying a generic mapping adjunct to core mapping (1) seeing ↔ knowing, viz. a power-adjunct

> WHERE V CORRESPONDS TO V*
> THERE power to V-y CORRESPONDS TO power to V*-y.

For the case of relations (to which we restricted attention in this paper):

(POWER) IF a relation Rxy in the SD
CORRESPONDS TO
a relation R*xy in the TD
THEN the power of x to stand in R to y in the SD
CORRESPONDS TO
the power of x to stand in R*-to y in the TD.

Instead, however, proponents of the introspective conception of the mind correlate the visual field with the personal space in which we keep things we when think of them (Section 6). This correlation has a fundamental, if perhaps well-hidden defect: It is no relevant source-target mapping; it fails to correlate an element of the visual SD with an element of the intellectual TD. For recall (from the end of Section 5) that in interpreting spatial cognition metaphors we use 'the mind' exclusively to stand for an element of these metaphors' source domain, namely, the personal space that the pretence scenario assigns to subjects. So the correlation of visual fields with minds correlates an element of the SD of vision cognition metaphors with an element of the source [!] domain of another conceptual metaphor, namely, the spatial cognition metaphor BEING THOUGHT OF AS BEING IN A PERSONAL SPACE. And while of course one conceptual metaphor's SD may, in principle, be another's TD, this is not the case here: All the conceptual metaphors at issue now have abstract TDs comprising related intellectual activities and achievements, and the mapping of visual fields to mind-spaces is from one concrete SD to another concrete SD which does not overlap with any of the intellectual TDs. So some sort of mistake must be involved in the particular SMT-style mapping operations that delivered this supposed SD-TD mapping.

To identify this mistake, consider again the representations from which the mapping was obtained (in Section 6), to repeat (for the reader's convenience):

(1) The visual field is the space in which the things are that the subject sees.

(2) The mind is the space in which the things are that the subject knows.

Knowledge representations may employ terms either literally or metaphorically. However, where we employ terms metaphorically in the representation of knowledge about the TD of a relevant conceptual metaphor, we may not simply assume that all the concepts employed by the representation stand for elements of the TD. Instead, we need to explicitly mark all metaphorical uses: Without further ado, we may then assign to the TD only those concepts that fall outside the scope of metaphorical use. Which elements of the TD are invoked by the metaphorically used expressions is something we need to make explicit by deriving their metaphorical interpretations. Only the concepts figuring in these interpretations can then be added to the TD stock.

Our representation (2) mixes literal with metaphorical uses of expressions: ‘subject’ and ‘knows’ are employed literally. But the phrase ‘the space in which things are’ is used metaphorically: We want to say that the mind is ‘the space in which things are’ when they are ‘kept in mind’; the ‘space’ at issue is the ‘mind’ in which they are then ‘kept’. The former clearly is a source-domain inference about the SD of the spatial cognition metaphor we use to interpret ‘S keeps X in mind’: When things are literally kept in a space or place, they are in that space or at that place. Conclusions of such reasoning may only be transferred to the TD upon analogical inference (Section 3). For applications in the TD, they hence have to be interpreted metaphorically, rather than literally. The talk in (2) of a ‘space in which things are’ is hence metaphorical. Marked accordingly, (2) becomes

(2’) The mind =def.: what, metaphorically speaking, we refer to as the space in which the things are that, literally speaking, the subject can know (viz. remember, when, metaphorically speaking, she keeps them in mind).

According to MAT (and ATT-Meta), metaphorically used expressions are to be interpreted by placing them into the ‘pretence cocoon’, where they are interpreted as true claims about the SD (not the TD!) of the relevant conceptual metaphor and further inferences are drawn. Their conclusions can then serve as premises of analogical inferences which use a restricted range of mappings; only the conclusions of such analogical inferences are then literally true of the TD and refer exclusively to elements of the TD (Section 3). The relevant inferences (Section 5) use the mapping (I) of the spatial relation ‘X is in the personal space (mind) of S $\leftrightarrow$ S thinks of X’ but do not map ‘the mind’ itself on anything. These analogical inferences thus lead to conclusions that no longer refer to ‘the mind’. We hence cannot assign the ‘mind-space’ of (2) to the TD, at any point of the interpretation process.

When we do so, nonetheless, this is arguably because we fail to distinguish between literal and metaphorical use of terms in the representation of TD knowledge from which we start out. Thus, we obtained mapping M above (Section 6) by treating ‘the space in which the things are’, in (2), as literally invoking a ‘space’ and the spatial relation ‘X is in Y’ (which they do when, and only when, the metaphorically used expressions are placed in the pretence-cocoon for further derivation of mappable SD conclusions) *and* thereby referring to TD elements (which they do not, at any point, as they do not get correlated with any TD element – only ‘X is in the space of S’ but not ‘X is in Y’ get mapped in metaphor interpretation).

The two mapping fallacies we have identified illustrate two ways in which metaphorical expressions can come to be interpreted overly literally: We may either overextend the underlying conceptual metaphor by adding to it mappings that are inconsistent with it, in the sense explained, and thus prevent default metaphorical interpretation. Or we may import elements of a conceptual metaphor’s concrete SD into an abstract TD. The former may happen as a result of analogy-maximising mapping strategies which we employ in problem solving but not in metaphor interpretation. The latter may result from such strategies when we fail to distinguish between literal and metaphorical uses of terms in representations of TD knowledge that are employed in alignment and mapping. Both fallacies occur in the derivation of the central tenets of introspective conceptions of the mind.

8. Conclusion and Future Research

We have thus obtained the outline of a debunking explanation of how those philosophically and culturally influential conceptions are obtained. By exposing fallacies in the inferences with which their central tenets are inferred, while letting us understand why we make them anyway, this explanation can help us resolve classical philosophical paradoxes which presuppose introspective conceptions of the mind, such as, e.g., 'arguments from illusion' and 'from hallucination' (Crane & French 2015, Fischer et al. 2015, Smith 2002). They also let us resolve paradoxes which arise from the clash of introspective conceptions with recent findings from social psychology (Bargh et al. 1996, Wilson 2002) and cognitive psychology (Gigerenzer et al. 2011, Kahneman 2011), which suggest that 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 of the sort introspective conceptions of the mind imply we have.

Further research is required to develop the proposed explanation: First, the new MAT version of the ATT-Meta approach remains to be computationally implemented by providing the ATT-Meta model with a comprehensive knowledge base that captures precisely the kind of world knowledge encoded in stereotypes (as initiated by Veale & Hao 2008) and can support rich interpretations of visual and spatial cognition metaphors. A computational implementation can be used to derive metaphorical interpretations for a wider range of expressions. Plausibility ratings for competing interpretations and paraphrase elicitation tasks (cp. Glucksberg & Haught 2006, Rubio-Fernandez et al 2015) can then be used to further examine the hypothesis that MAT (but not, say, SMT) captures a default strategy for metaphor interpretation by testing predictions generated by the computational model. Plausibility-ratings for solutions to problems presented in different metaphorical and literal frames can then be used to examine the other half of the differential processing hypothesis and the conditions under which thinkers use linguistically realised conceptual metaphors for analogical reasoning in problem-solving (Jia & Smith 2013, Keefer et al. 2014). Case-studies on philosophical texts can finally study the extent to which potential conclusions of such reasoning are accepted without, and presupposed in further philosophical argument (cp. Cappelen 2012, Fischer 2011), so that our warrant for maintaining introspective conceptions of the mind is dependent upon what spontaneous inferences are made for obtaining their key tenets. According to the proposed explanation, these inferences are fallacious.

References

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

Barnden, J. 1997. Consciousness and common-sense metaphors of mind. In S. O'Nuallain, P. McKevitt & E. Mac Aogain (eds.), *Two Sciences of Mind: Readings in Cognitive Science and Consciousness* (pp.311-340). John Benjamins.

Barnden, J.A. 2001: Uncertainty and conflict handling in the ATT-Meta context-based system for metaphorical reasoning. *Lecture Notes in Artificial Intelligence*, 2116 (pp.15–29). Berlin: Springer.

Barnden, J.A. 2008: Metaphor and artificial intelligence: Why they matter to each other. In R.W. Gibbs (ed.), *The Cambridge Handbook of Metaphor and Thought* (pp.311–338). Cambridge: CUP.

Barnden, J.A. 2015: Open-ended elaborations in creative metaphor. In T.R. Besold, M Schorlemmer and A. Smaill (eds.), *Computational Creativity Research: Towards Creative Machines* (pp.217–242). Berlin: Springer

Barnden, J.A. 2016. Mixed metaphor: Its depth, its breadth, and a pretence-based approach. In R.W. Gibbs (ed.), *Mixing Metaphor* (pp. 75-111). Amsterdam: John Benjamins

Barnden, J.A., Glasbey, S.R., Lee, M.G. & Wallington, A.M. 2002: Reasoning in metaphor understanding: The ATT-Meta approach and system. *Procs. 19th International Conference on Computational Linguistics*(COLING 2002), pp.1188-1193. San Francisco: Morgan Kaufman.

Barsalou, L. 2007: Grounded cognition. *Annual Review of Psychology* 59, 617-645

Bartha, P. 2010: *By Parallel Reasoning*. Oxford: Oxford University Press

Bartha, P. 2013: Analogy and Analogical Reasoning. *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2013/entries/reasoning-analogy/>

Bennett, M.R. and Hacker, P.M.S. 2003: *Philosophical Foundations of Neuroscience*. Oxford: Balckwell

Bortfeld, H. 1998: Across-linguistic analysis of idiom comprehension by native and non-native speakers. *Dissertation Abstracts International: Section-B: Sciences and Engineering* 59, 0432.

Bortfeld, H. and McGlone, M.S. 2001: The continuum of metaphor processing. *Metaphor and Symbol* 16, 75-86

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

Cappelen, H. 2012: *Philosophy without Intuitions*. Oxford: OUP

Crane, T. and French, C. 2015: The problem of perception. In N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Winter 2015 version, URL = <http://plato.stanford.edu/entries/perception-problem/>

Danesi, M. 1990: Thinking is seeing: Visual metaphors and the nature of abstract thought. *Semiotica, 80*, 221–238.

Falkenhainer, B., Forbus, K.D. and Gentner, D. 1989: The structure-mapping engine: algorithm and examples. *Artificial Intelligence 41*, 1-63

Fein, O., Yeari, M., and Giora, R. 2015: On the priority of salience-based interpretations: the case of sarcastic irony. *Intercultural Pragmatics* 12, 1-32

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

Fischer, E. 2011: *Philosophical Delusion and its Therapy*. New York: Routledge

Fischer, E. 2014: Philosophical intuitions, heuristics, and metaphors. *Synthese*, 191: 569-606

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

Fischer, E. and Engelhardt, P.E. 2016: Intuitions' linguistic sources: Stereotypes, intuitions, and illusions. *Mind and Language* 31, 67-103.

Fischer, E., Engelhardt, P.E. and Herbelot, A. 2015: Intuitions and illusions. From explanation and experiment to assessment. In: E. Fischer and J. Collins (eds.), *Experimental Philosophy, Rationalism and Naturalism* (pp. 259-292). London: Routledge

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

Forbus, K. D., Ferguson, R. W., Lovett, A. and Gentner, D. 2016. Extending SME to handle large-scale cognitive modelling. *Cognitive Science*. doi:10.1111/cogs.12377

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

Gentner, D. and Bowdle, B. 2008: Metaphor as structure-mapping. In R. Gibbs (ed), *The Cambridge Handbook of Metaphor and Thought*. New York: CUP, pp. 109-128

Genter, D. and Forbus, K.D. 2011: Computational models of analogy. *WIREs Cognitive Science 2*, 266-276

Gentner, D., Imai, M. and Boroditsky, L. 2002: As time goes by: evidence for two systems in processing space-time metaphors. *Language and Cognitive Processes 17*, 537-565

Gentner, D. and Markman, A. B. 1997: Structure mapping in analogy and similarity. *American Psychologist* 52, 45-56.

Gentner, D. and Markman, A.B. 2005: Defining structural similarity. *Journal of Cognitive Science* 6: 1-20

Gentner, D., Ratterman, M. and Forbus, K. 1993: The roles of similarity in transfer: separating retrievability from inferential soundness. *Cognitive Psychology* 25, 527-575

Gerrig, R. and Healy, A. 1983: Dual processes in metaphor understanding: Comprehension and appreciation. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 9, 667–675.

Gibbs, R.W. Jr. 2006: Metaphor interpretation as embodied simulation. *Mind and Language* 21: 434-458

Gibbs, R.W. Jr. 2011: Evaluating Conceptual Metaphor Theory. *Discourse Processes* 48, 529-562

Gibbs, R.W. Jr. & O'Brien, J.E. 1990: Idioms and mental imagery: The metaphorical motivation for idiomatic meaning. *Cognition* 36, 35 68

Gigerenzer, G., Hertwig, R., & Pachur, T. (eds.) 2011: *Heuristics: The foundations of adaptive behavior*. New York: Oxford University Press.

Giora, R. 2003: *On Our Mind. Salience, Context, and Figurative Language*. Oxford: OUP

Giora, R. and Fein, O. 1999: On understanding familiar and less-familiar figurative language. *Journal of Pragmatics* 31, 1601-1618

Giora, R., Givoni, S. and Fein, O. 2015. Defaultness reigns: The case of sarcasm. *Metaphor and Symbol* 30, 290-313

Giora, R., Raphaely, M., Fein, O., and Livnat, E. 2014: Resonating with contextually inappropriate interpretations in production: The case of irony. *Cognitive Linguistics* 25, 443-455

Glucksberg, S. and Haught, C. 2006: On the relation between metaphor and simile: When comparison fails. *Mind and Language* 21: 360–378.

Goatly**,** A. 1997: The language of metaphors. London: Routledge

Goschler, J. 2005: Embodiment and body metaphors. *Metaphorik.de*, *9*, 33–52.

Grady, J.E. 1997: Theories are buildings revisited. *Cognitive Linguistics* 8, 267-290

Gresko, J. 2015: *Love after 5th Avenue*. Lulu Press.

Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. 2009: Activating event knowledge. *Cognition* 111, 151-167

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

Herrero Ruiz, J. 2003: The emeralds of your face. Metaphors and metonymy in some expressions. *Cuadernos de investigación filológica* 29-30, 135-157

Hobbs, J.R. 1992: Metaphor and abduction. In A. Ortony, J. Slack and O. Stock (eds.), *Communication from an Artificial Intelligence Perspective: Theoretical and Applied Issues* (pp. 35–58). Berlin: Springer

Hodgetts, C.J., Hahn, U. and Chater, N. 2009: Transformation and alignment in similarity. *Cognition*, 113: 62–79

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

Holyoak, K.J. and Thagard, P. 1989: Analogical mapping by constraint satisfaction. *Cognitive Science*, 13: 295–355

Hummel, J. and Holyoak, K.J. 1997: Distributed representation of structure: A theory of analogical access and mapping. *Psychological Review*, 104: 427–466

Jaszczolt, K. M. 2011: Default meanings, salient meanings, and automatic processing. In K.M. Jaszczolt and K. Allan (eds.), *Salience and defaults in utterance processing* (pp. 11–33). Berlin: De Gruyter

Jia, L. and Smith, E.R. 2013: Distance makes the metaphor grow stronger. *Journal of Experimental Social Psychology* 49, 492-497

Kahneman, D. 2011: *Thinking Fast and Slow*. London: Allen Lane

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

Koivisto-Alanko, P. & Tissari, H. 2006: Sense and sensibility: rational thought vs emotion in in metaphorical language. In A. Stefanowitsch and S.T. Gries, *Corpus-based Approaches to Metaphor and Metonymy* (pp. 191-213). New York: De Gruyter

Kövecses, Z. 2010: *Metaphor: A practical introduction*, 2nd ed. New York, NY: Oxford University Press.

Lakoff, G. 1994: Master metaphor list. University of California. http://meta-guide.com/data-processing/computational-metaphorics/master-metaphor-list/

Lakoff, G. and Johnson, M. 1980: *Metaphors We Live By*. Chicago: Chicago UP

Lakoff, G. and Johnson, M. 1999: *Philosophy in the Flesh*. New York: Basic Books

Landau, M. J., Keefer, L. A. and Rothschild, Z. K. 2014: Epistemic motives moderate the effect of metaphoric framing on attitudes. *Journal of Experimental Social Psychology*, 53: 125–138

Lassaline, M.E. 1996: Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22, 754-770

Lee, M.G. and Barnden, J.A. 2001: Reasoning about mixed metaphors with an implemented AI system. *Metaphor and Symbol*, 16: 29–42

Levinson, S.C. 1983: *Pragmatics*. Cambridge: Cambridge University Press

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

MacArthur, F., Krennmayr, T. & Littlemore, J. 2015: How Basic Is "UNDERSTANDING IS SEEING" When Reasoning About Knowledge? *Metaphor and Symbol*, 30: 184-217

McDonald, P. 2003: *History of the Concept of Mind*. Aldershot: Ashgate

Markman, A. 1997: Constraints on analogical inference. *Cognitive Science 21*, 373-418

McKoon, G., & Ratcliff, R. 1989: Semantic association and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15, 326-338.

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

McRae, K., Hare, M., Elman, J. L., & Ferretti, T. 2005: A basis for generating expectancies for verbs and nouns. *Memory & Cognition* 33, 1174–1184.

McRae, K. and Jones, M. 2013: Semantic memory. In D. Reisberg (ed.), *Oxford Handbook of Cognitive Psychology*, Oxford: OUP

Metusalem, R., Kutas, M., Urbach, T.P., Hare, M., McRae, K. and Elman, J.L. 2012: Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language* 66, 545–567

Narayanan, S. 1999: Moving right along: A computational model of metaphoric reasoning about events. *Proceedings of the National Conference on Artificial Intelligence (AAAI 99)* (pp.121–128). AAAI Press.

Neely, J.H. 1991: Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner and G. Humphreys (eds.), *Basic Processes in Reading: Visual Word Recognition*. Hillsdale: Erlbaum, pp. 264-336

Peleg, O. and Giora, R. 2011: Salient meanings: The whens and wheres. In K.M. Jaszczolt and K. Allan (eds.) *Salience and Defaults in Utterance Processing* (32 – 52). Berlin: de Gruyter.

Peleg, O., Giora, R. and Fein, O. 2004: Contextual strength: The whens and hows of context effects. In I. Noveck and D. Sperber (eds.), *Experimental Pragmatics* (pp. 172-186). Basingstoke: Palgrave

Rorty, R. 1980: *Philosophy and the Mirror of Nature*. Oxford: Blackwell

Rubio-Fernández, P., Wearing, C. & Carston, R. 2015: Metaphor and Hyperbole: Testing the Continuity Hypothesis. *Metaphor and Symbol* 30: 24-40

Rumelhart, D.E. 1980. Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, and W. Brewer (eds.), *Theoretical Issues in Reading Comprehension* (pp. 33-58). Hillsdale, N.J.: Erlbaum.

Searle, J. 1993: Metaphor. In A. Ortony (ed.), *Metaphor and Thought, 2nd ed.*, (pp. 83-111). Cambridge: CUP

Smith, A.D. 2002: *The Problem of Perception*. Cambridge, Mass: Harvard UP

Steen, G.J. 2011. The contemporary theory of metaphor – now new and improved. *Review of Cognitive Linguistics* 9, 26-64.

Steen, G.J., Dorst, A.G., Herrmann, J.B., Kaal, A., Krennmayr, T. and Pasma, T. (2010). *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. Philadelphia: John Benjamins.

Sweetser, E. 1991: *From Etymology to Pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.

Thibodeau, P.H. and Boroditsky, L. 2011. Metaphors we think with: The role of metaphor in reasoning. *PLoS ONE* 6, e16782.

Thibodeau, P.H. and Boroditsky, L. 2013. Natural language metaphors covertly influence reasoning. *PLoS ONE* 8, e52961.

Thibodeau, P.H. and Boroditsky, L. 2015. Measuring effects of metaphor in a dynamic opinion landscape. *PLoS ONE* 10, e0133939.

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

Traugott, E.C. and Dasher, R.B. 2005. *Regularity in Semantic Change*. Cambridge: CUP.

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

Vandeleene, A., Dodeigne, J., Heyvaert, P., Legein, T., Perrez, J., and Reuchamps, M. In prep. The mediating role of political knowledge on the impact of metaphors: Two experimental studies on political preferences for basic income and federalism in Belgium. Presented at *6th Conference of the Scandinavian Association for Language and Cognition*, Lund, Sweden, April 19-12, 2017.

Veale, T. and Hao, Y. 2008. Enriching WordNet with folk knowledge and stereotypes. In *Proceedings of the 4th Global WordNet Conference, Szeged, Hungary*.

Wallington, A.M. 2010: Systematicity in metaphor and the use of invariant mappings. In G. Low, Z. Todd, A. Deignan & L. Cameron (eds), *Selected Papers from "Researching and Applying Metaphor in the Real World"* (pp.209-244). Amsterdam: Benjamins.

Walton K. 2004: Fiction and non-fiction. In E. John & D.M. Lopes (eds), *Philosophy of Literature – Contemporary and Classic Readings* (pp.136–143). Oxford: Blackwell.

Weinberg. J. 2015: Humans as instruments, on the inevitability of experimental philosophy. In: E. Fischer and J. Collins (eds.): *Method, Rationalism, and Naturalism* (pp. 171-187). London: Routledge

Wilson, T.D. 2002. *Strangers to Ourselves. Discovering the adaptive unconscious*. Cambridge, Mass.: Harvard UP

Wittgenstein, L. 1933/2005: *The Big Typescript: TS213*. Oxford: Blackwell

Wolff, P. and D. Gentner. 2011: Structure-mapping in metaphor comprehension. *Cognitive Science* 35: 1456–488

Tables

Table 1. Spatial cognition metaphors in German with English translation

in den Sinn kommen	come to mind
durch den Sinn fahren	cross (through) the mind
durch den Kopf gehen	go through the mind/head
im Gedächtnis behalten	keep in mind
in Erinnerung rufen	call to mind
sich aus dem Kopf schlagen	put (literally: hit) out of one's mind
aus dem Gedächtnis verbannen	banish from the mind

Table 2. A CWSG inference with transcendent mapping

	SD premise	**Operation**	**TD conclusion**
1	S looks at X	Substitution: core mapping (2) (Looking at ↔ Thinking about)	S thinks about X
2	(1) Implies (3-4)	Substitution: identical	(1) Implies (3-4)
3	X before Y	Generation[27]	X before Y
4	Y=eyes(S)	Substitution: mapping N	Y=understanding(S)

[27] The present analysis explores what happens in CWSG inference with narrow vision-cognition metaphors complemented by N and M. Since these narrow conceptual metaphors don't map spatial relations, 'X is before Y' is left over in the candidate conclusion, and thus generated in the TD, once all substitutions have been made. Where more comprehensive mappings are used and representations of TD knowledge are taken to invoke suitable correlates, CWSG inferences use substitution, instead, to arrive at the same conclusions.