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

Eugen Fischer[1]

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

1 INTRODUCTION

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

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

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

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

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

2 TWO STRATEGIES: FULL-BLOODED AND RESTRICTED ANALOGICAL REASONING

Our argument will rely on the distinction between two strategies:

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

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

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

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

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

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

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

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

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

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

(1) S sees x → S knows x

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

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

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

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

Table 1. An elementary CWS inference

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

Other ECWS inferences yield, e.g.:

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

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

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

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

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

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

(AI) S cannot get to know X.

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

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

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

3 METAPHORICAL MINDS

As we will now see, introspective conceptions of the mind essentially rely on rather more 'full-blooded' analogical reasoning that (a) involves copying with substitution *and generation* (full CWSG) and (b) invokes both mappings constitutive of visual metaphors and further mappings that 'transcend' these metaphors. While the terminology varies slightly, seminal early modern texts work with the twin mappings (see, e.g., Fischer [34] on Locke [35]):

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

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

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

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

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

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

Table 2. A CWSG inference with transcendent mapping

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

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

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 things with our understanding, *in* our mind.

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

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

Relevant inference from P_1 yields

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

This has a literal interpretation (which is true: when I think – or do anything else, for that matter – something or other will be in front of my eyes, and sometimes I even think about the very things then in front of me). Crucially, it also has a metaphorical interpretation motivated by the visual metaphor: When something is before my eyes, it is typically easy for me to notice (get to see). Stereotypical inference therefore furnishes the premise for an ECWS inference to the conclusion that it is easy for me to get to know or understand. This yields this interpretation of C_1*:

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

– perhaps unduly optimistic and not idiomatic, but intelligible.

Similarly, analogical inference without M leads from P_2 to

C_2* When we think about things, things are within our ken.

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

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

Since none of the elements P_3 refers to are mapped by the conceptual metaphors at issue, analogical inferences with these metaphors cannot be directly made from this premise. However, P_3 itself employs phrases which have stereotypical implications in the source domain of vision: When things are before our eyes, it is easy to see them, and when things are in our visual field, it is at any rate possible for us to see them. ECWS inferences lead from the conclusions of the corresponding stereotypical inferences to an undeniable conclusion:

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

Finally, analogical inference with visual metaphors but without M and N does not take us much beyond P_4: Since "perceive", explained by the *OED* as 'to apprehend with the mind or senses', stands for an epistemic relation that can obtain in both the SD of seeing and the TD of cognition, it initially gets mapped onto, and substituted by, itself. We thus obtain:

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

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

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

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

To sum up: Analogical reasoning with visual cognition metaphors only gets us from SD truisms (like P_1 to P_4) to the conclusions (C_1 to C_4) constitutive of the introspective conception of the mind, if we make use of further mapping (like M and N) which 'transcend' those metaphors.

4 ILLUSIONS OF SENSE

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

Our first set of conclusions, C_1 to C_4, is a case in point. In contrast with their starred counterparts, they lack metaphorical interpretations motivated by visual metaphors. They all employ at least one of two phrases we obtain when applying N and M to source-domain truisms: "before our understanding" and "in our mind". Neither has a metaphorical interpretation motivated by visual cognition metaphors: In contrast with the source-domain expression "x is before our eyes" from which it is obtained, "x is before our understanding" has no stereotypical or semantic implications in the visual SD. Hence there is nothing for visual cognition metaphors to map, and our default comprehension strategy of making ECWS inferences with mappings constitutive

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

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

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

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

(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.

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

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

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

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

5 EXPLAING THE TRANSCENDENT MAPPINGS

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

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

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

P_5 When we look at something, we use our eyes.

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

Mapping N: eyes → understanding

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

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

Parallel connectivity yields N in analogical reasoning from premises like P_5, with the core mappings of different related conceptual metaphors: 'Thinking-about as Looking-at', 'Understanding as Seeing', etc. Like many action- and event-nouns [30], all these verbs are associated with quite complex stereotypes known as 'generalised situation schemas' [16, 17]. These are made up of typical features of the action or event that the verb refers to, of the agents performing the action, and of the 'patients' on which it is performed. These features crucially include instruments typically used in performing the action [43]. The strength of stereotypical association is commonly measured through the 'cloze probability' or frequency with which the relevant concept is used to complete sentences such as:

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

The most frequent responses are (1) 'needle' and (2) 'the police' or 'cops' [17]. And while the cloze frequencies for (3) and (4) have not yet been systematically elicited, readers will have little trouble completing them with (3) 'eyes' and (4) 'brains' or 'minds', 'wits', 'reason', 'intelligence' – early moderns would have said our 'intellect' or 'understanding'. Arguably, just as 'sewing' is associated with the subject-property 'uses a needle', 'looking at' is associated with 'uses his eyes', and 'thinking about' with 'uses his brain / mind/ reason / understanding'.

When we encounter or use a verb, all the concepts belonging to the associated generalised situation schema are activated – irrespective of contextual relevance, and the more swiftly and strongly, the stronger the association is [44]. The more strongly a concept is activated, the more likely it will be used in various cognitive processes. If the subject is engaged in analogical reasoning, the concept is hence more likely to be mapped or generated. Where an action or event designated by a source-domain verb gets mapped onto a target-domain concept, all key elements of the situation schema associated with the verb are hence likely to be mapped or generated. Where the schema associated with the TD verb contains an element that stands in the same relation (say, the instrument-relation) to the TD action as the SD associate to the SD action, the SD associate will be mapped onto the TD associate – regardless of whether that relation actually figures in the premise. Thus 'eyes' get mapped onto 'mind' or 'understanding' even in inferences from premises in which the instrument-relation does not figure, like (P_1) 'When we look at things, things are before our eyes.' Enforcing the constraint of 1-on-1 mapping in reasoning that also employs mapping M, of 'visual field' onto 'mind', then leads to the preference of 'understanding' over 'mind' we can observe in early modern texts (cp. [34]).

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

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

(SI_1) When S keeps something x in a space (belonging to him), then X continues to be in the space belonging to S.

A mapping of this temporal relation onto the TD relation 'S continues to think of X' can be generated from the core Mapping R through ECWS inferences (what ATT-Meta calls 'vehicle neutral mapping adjuncts'). Analogical inference with this further mapping takes us from the consequent of (SI_1) to

(AI) S continues to think of X.

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

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

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

(SI_2) When something is at the forefront of my visual field, I cannot help looking at it.
(SI_3) When something is at the back of my visual field, I don’t look at it but am aware of it.

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

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

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

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

Once the crucial mistake of treating ‘the mind’ as a TD element has been made, standard mapping principles have us map ‘visual field’ onto it: In a first step, SD elements get mapped onto the TD elements deemed most similar to them (Section 2). Through post-inference replacements in antecedents of inferences like (SI_1) to (SI_3), the ‘mind’ appears to be credited will all the abstract features (properties and relations) of delineated spaces (in which things can be kept, etc.) and, more specifically, visual fields (which have forefronts and backs, i.e. depth). Through such apparent attributions, ‘visual fields’ and ‘minds’ come to be deemed similar enough to get mapped in the first stage of mapping. The presently relevant premises P_1 to P_4 do not provide any other relata for ‘x is in y’, so the mapping does not fall foul of structural constraints, in the second stage.

6 CONCLUSIONS

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

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

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[19] J.A. Barnden. Creative metaphor and metaphorical creativity. In T.R. Besold, M. Schorlemmer & A. Smaill (eds.): *Computational Creativity Research: Towards Creative Machines* (pp. 217-242). New York: Springer (2015).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[2] For helpful comments on a previous draft, I am indebted to John Barnden and an anonymous referee.