

# Analogical insight: toward unifying categorization and analogy

Eric Dietrich

Received: 6 July 2009 / Accepted: 21 June 2010  
© Marta Olivetti Belardinelli and Springer-Verlag 2010

**Abstract** The purpose of this paper is to present two kinds of analogical representational change, both occurring early in the analogy-making process, and then, using these two kinds of change, to present a model unifying one sort of analogy-making and categorization. The proposed unification rests on three key claims: (1) a certain type of rapid representational abstraction is crucial to making the relevant analogies (this is the first kind of representational change; a computer model is presented that demonstrates this kind of abstraction), (2) rapid abstractions are induced by retrieval across large psychological distances, and (3) both categorizations and analogies supply understandings of perceptual input via *construing*, which is a proposed type of categorization (this is the second kind of representational change). It is *construing* that finalizes the unification.

**Keywords** Analogy · Abstraction · Categorization · Concepts · Insight · Knowledge change · Knowledge representation · Mapping · Psychological distance · Representational change · Retrieving · Understanding

## Introduction: analogy and representational change

In their 2003 *Cognitive Science* paper, psychologists Ramscar and Yarlett say:

[There is an] increasing amount of evidence that the mechanisms underlying analogy and similarity-based transfer play a crucial role in *all* conceptual thought, and critically in categorization. From a knowledge-representation and processing point of view, explaining both ordinary conceptual processes and analogy are very similar tasks. (Ramscar and Yarlett 2003, pp. 67–68)

This quote nicely sums up one main thesis of this paper which is that analogy and categorization are deeply similar. A proposed approach to unifying the two will be presented here.<sup>1</sup> This approach requires postulating two kinds of representational change, one which, it will be argued, happens right at the time of analogical retrieval (of the relevant kind), and the other immediately thereafter. This takes us to the paper's other main thesis. To introduce it, consider a second Ramscar and Yarlett quote:

[Our mapping engine, M-SME, and Gentner's engine, SME] succeed in ordering the candidate bases appropriately in terms of their underlying structural commonalities with the target analog in the absence

<sup>1</sup> Others also have the goal of unifying analogy and categorization; see, e.g., Atkins (2004) and Kuehne et al. (2000). Several researchers are interested in unifying analogy with other cognitive processes. Three examples are unifying analogy with reasoning [e.g., Hummel and Holyoak 2003 (in this paper, inference is added to the capacities of their analogy model, LISA 1997); Hummel and Choplin 2000; Kokinov and Petrov 2001], unifying analogy with perception (e.g., Mitchell 1993), and unifying analogy with schema induction and learning (e.g., Hummel and Holyoak 1997, 2003). In several of these projects, analogy is not so much unified with some other cognitive process, rather that cognitive process is shown to require or depend on analogy (usually analogical mapping) in order to work properly. Good examples of such an approach can be found in Dumas et al. (2008) and Hummel and Holyoak (2003).

E. Dietrich (✉)  
Department of Philosophy, Binghamton University,  
Binghamton, NY, USA  
e-mail: dietrich@binghamton.edu

of any “semantic” information. [This may be due to] the fact that once representations that have appropriate structural commonalities have been retrieved, a structure mapping algorithm will, by definition, be able to determine which structures match the best. This consideration leads to a problem with *all* current models of analogy (including [Ramscar’s and Yarlett’s]): *the representations used in the simulations reported here are static* (they encapsulate knowledge in a “canned” format). (Ramscar and Yarlett 2003, p. 67; emphases added.)

Ramscar and Yarlett go on to point out that it is questionable whether human knowledge used during analogy making is as canned and neatly packaged as assumed in many simulations of analogy (p. 67). They then say:

Although evidence suggests that analogical reasoning is incremental (Keane 1996, 1997) with base representations [those “retrieved” from long-term memory and forming the basis of the analogy] being built up dynamically as reasoning proceeds, the processes underlying representation building are poorly understood. *How are these representations built up?* (2003, p. 67; emphasis added.)

The second main purpose of this paper is to propose and explore one answer to their last question. However, the kind of representational changes proposed here are not incremental and do not involve building representations up over time. This is not to deny that incremental, representation construction occurs in analogical reasoning, the evidence for such construction is strong (e.g., Keane and Brayshaw 1988). It is, rather, to suggest that (at least) two other kinds of change are also present; indeed, required. These changes occur not in the reasoning phase, but right at the time the analogy is made. The first of these two is a species of abstraction and is quite rapid (on the order of milliseconds). I will refer to it simply as *rapid abstraction*. The second is, usually, a much milder sort of change and forms the basis of unifying analogy and categorization. It is called *construing*, and it is a type of categorization.

Here is another view on Ramscar’s and Yarlett’s key question. Since our categories are the basis for how we understand our environment, relating analogy to categorization has as its immediate consequence that analogy is also a way of producing understanding. This is important, for sometimes analogy researchers talk as if the freshness of an experience of analogy resides solely in seeing that something is like something else—seeing that the atom is like a solar system, that heat is like flowing water, that paint brushes work like pumps, or that electricity is like a teeming crowd. But analogy is more than this. Analogy is not just seeing that the atom is like a solar system; rather, it is seeing something

*new* about the atom, an observation enabled by ‘looking’ at atoms from the perspective of one’s understanding of solar systems. Analogy and categorization as well are both at root dynamical constructions of an *understanding* of a perceived object; in the case of analogy, this understanding is usually *new* (at least to the individual having the analogy). The relevant understanding is due, in part, to a specific kind of knowledge *creation* which requires the two sorts of representational change proposed here.<sup>2</sup> Our question then is this: How can an analogy provide new knowledge and new understanding? This is another version of Ramscar’s and Yarlett’s question, above. Cognitive scientists, of course, have produced many different answers to this question as they’ve struggled to understand analogy and its place in cognition. The answer suggested here is intended to be an addition to those already out there.

The thesis that analogy *requires* rapid representational change contrasts with one dominant view of how analogy works: “A good analogy both *reveals* common structure between two situations and suggests *further inferences*” (Gentner and Colhoun 2010, emphases added). Revealed structure exists ahead of time, so Gentner’s view is that the relevant representational structure (and hence knowledge) associated with analogy-making exists ahead of time—before the analogy is completed, or even begun—and that only after an analogy is made do further inferences kick in and possibly alter or suggest structural alterations and refinements (for the seminal work on this, see Gentner 1983; see also Gentner and Wolff 2000). As important as this view has been and continues to be in analogy research, it has also contributed to the problem of static representations Ramscar and Yarlett flag above. The proposal here does not completely deny Gentner’s point, which is widely held—in some cases, similarity of structure no doubt exists ahead of time—but instead adds to it: to make certain analogies, rapid abstraction is required.

To unify analogy and categorization, three ingredients are required: (1) a specification of rapid abstraction and an argument that it in fact needs to be added to the kinds of representational change already studied in analogy research, (2) showing that building such abstractions is not as computationally expensive as might be thought; this is accomplished by adding in Liberman’s and Trope’s work on abstraction and psychological distance (Liberman and Trope 2008), and (3) the *construing* model of categorization, which will give us the actual unification. The first of these is done in “[Semantically distant analogies and rapid abstraction](#)” and “[Rapid abstraction in STRANG](#)”, the second

<sup>2</sup> Others who have noted the importance to analogy, and to conceptual combination in general, of different kinds of knowledge creation include: Day and Goldstone (2009), Dixon and Dohn (2003), Indurkha (1992, 1998), Wisniewski (1997).

in “[Abstraction and psychological distance](#)”, and the last in “[Construing and categorization](#)” and “[Construal-based analogy: unifying analogy and categorization via construing](#)”. An example of the whole is shown in “[A simple example](#)”. “[Conclusion](#)” concludes. In detail: in “[Semantically distant analogies and rapid abstraction](#)”, needed background is presented as well as the case that rapid abstraction is required at the time of analogy-making. In “[Rapid abstraction in STRANG](#)”, a computer model, STRANG, is used to further demonstrate rapid abstraction. “[Abstraction and psychological distance](#)” argues that Liberman’s and Trope’s work on psychological distance shows that the computational expense of such abstraction is low. “[Construing and categorization](#)” briefly presents the construing model (Kurtz and Dietrich 2006). “[Construal-based analogy: unifying analogy and categorization via construing](#)” shows how construing and the sort of analogy-making considered here have to be combined. This gives us our unification. All of this is brought together in “[A simple example](#)”, in an example. The result is new additions to the kinds of knowledge change associated with analogy-making and analogical reasoning and a unification of one type of categorization with a kind of analogy.

A few words of clarification. This paper does not offer a new theory of the whole complex analogy-making and reasoning process. For one thing, only one small part of the whole process is considered here, though it is a crucial part. This is the part involving rapid abstraction and construing, and it occurs early in the analogy-making process. Secondly, on many issues in analogy research, this paper is rather orthodox. For example, the distinction between retrieving potential analogues from memory (also called “access”) and then mapping one or more of them to items already in working memory is preserved. The importance of mapping is also preserved. Preserved also are the notions of structured representations (representations with connected, multi-place predicates) and their role in analogical mapping. In fact, the first kind of change, rapid abstraction, could be modeled within some well-known analogical retrieval engines such as MAC/FAC (Forbus et al. 1995) by adding the relevant representational change algorithms.<sup>3</sup> However, the paper is not orthodox in its claims about the role of rapid representational change in making certain

<sup>3</sup> For example, in the case of MAC/FAC (“many are called, but few are chosen”), rapid abstraction could be added between the MAC and FAC stages; i.e., between the computationally cheap matcher/retriever (MAC) and the more expensive and sophisticated final filterer and selector (FAC—which is basically SME, the Structure Mapping Engine, see Falkenhainer et al. 1989). Alternatively and more efficiently (especially given Liberman’s and Trope’s experimental data, see “[Abstraction and psychological distance](#)”, below), with appropriate modifications, rapid abstraction could just be incorporated into the MAC stage.

analogies, nor in its advocating construing as a part of the analogy-making process.

### Semantically distant analogies and rapid abstraction

First, I explain the types of analogy focused on here. Then, I discuss some details of the relevant kind of analogy-making process and argue that this kind needs rapid analogical abstraction.

#### Insightful analogies

The type of analogy of interest here is spontaneous analogical reminding occurring over a large semantical distance; such analogies occur often in humans. There is some dispute about how often. I follow Dunbar in this matter: they occur frequently (see, e.g., Dunbar 2001). Dunbar locates the explanation of this frequency in the fact that such analogies are *generated*; accordingly, I will be focusing also on the generation of such analogies. It does not follow from this assumption, nor am I suggesting that insightful analogies which produce robust, useful knowledge (like scientific discoveries) occur frequently. Those seem to be somewhat rare. But spontaneous analogical reminders across large distances are rather common. An example: A colleague was cross-country skiing with his wife. They paused to rest and drink some water. Though it was cold, they were quite warm and his wife, being mindful of hypothermia caused by overheating, took off one glove to cool down. She then quipped: “My hand is like a dog’s tongue when he’s panting in the summer.”

Another, more important case, comes from the history of science; it is famous in analogical circles: Ernest Rutherford’s 1909 experiments on atomic structure and his discovery of the atomic nucleus. Rutherford bombarded thin gold foils (about 400 atoms thick) with alpha particles (helium nuclei) and then tracked where the particles went.<sup>4</sup> Assuming the then dominant Thomson model of the atom (sometimes called the “plum pudding model”), Rutherford predicted that the vast majority of alpha particles targeted at a foil would pass more or less straight through it, perhaps with some being mildly deflected in their paths through the foil. But that is not what happened. Rutherford noted that about one out of every 8,000 alpha particles exhibited strongly deflected trajectories, basically “bouncing back” in the direction from which they had come, a phenomenon that completely took him by surprise, and which was incompatible with the Thomson model. He is reported to

<sup>4</sup> The actual experiments were run by Hans Geiger and Ernest Marsden in 1909, and they discovered the scattering alpha particles.

have said “It was almost as incredible as if you fired a fifteen-inch shell at a piece of tissue paper and it came back and hit you.” The strongly deflected alpha particles roughly followed a hyperbolic path relative to the gold atoms (I say “roughly” since the particles were “bouncing off” the atoms’ nuclei). Comets that do not orbit the sun but make one pass and then are never heard from again, also travel hyperbolic paths. There is evidence that the seemingly hyperbolic paths of the particles reminded Rutherford of comets and their paths, and that he used this analogy to develop an explanation of the results of the gold foil experiments (see Gentner and Wolff 2000, and Wilson 1983). Rutherford’s *extended* analogical reasoning (much more than just the comet analogy) suggested that most of the mass of the atom was in its center, just the opposite of the Thomson model. His completed analogically derived model (not the experimental data behind it) doomed the reigning plum-pudding conception of the atom and led eventually to the better Bohr model of the atom.

The dog-tongue analogy, Rutherford’s comet analogy, and his tissue paper analogy well exhibit the sort of representational change that I am interested in. Consider the comet analogy. Deflected alpha particle trajectories are by no means orbits, nor are they closely similar to hyperbolic comet paths, at the very least because gravity is not causing the deflection. Removing gravity as a cause as well as the notion of orbits are key changes to the relevant representations. Such changes had to occur in order to forge the analogy, furthermore, they had to occur quickly. We will return to this example in “A simple example”, where we will consider further details.

#### Background on analogical representational change

Representational change of many types is an important research topic throughout much of cognitive science. Analogical research is an exemplar of this. Almost all psychological theories of analogy propose some sort of representational change associated with making an analogy or with using it to make inferences or new knowledge. One good example is *re-representation* whereby—e.g., in the solar system/atomic structure analogy—“revolving around” and “orbiting” are decomposed to (say) “circumnavigating” (c.f., Ramscar and Yarlett 2003, p. 43), or “lift” and “raise” are decomposed to “cause to rise” (Hummel and Holyoak call this *chunking*; see their 1997).

Classical Structure Mapping Theory (SMT; Gentner 1983) does not allow for representational change during the early part of analogy-making. But Gentner and Wolff (2000) provided an extensive catalogue of representational changes that work within SMT. All of these are clearly important, but all occur after an analogy is made.

LISA (Hummel and Holyoak 1997, 2003) can generate inferences, induce schemas, generalize, chunk, and learn. These are all important and interesting representational changes. Again, these occur post-memory access and analogical mapping, as LISA in many ways depends on structure mapping (2003, p. 228). However, LISA can also map  $m$ -place predicates to  $n$ -place ones,  $n \neq m$ . In certain cases, this sort of capacity obviates the need for representational changes to predicate relations. Still, both the  $n$ -place and  $m$ -predicates exist ahead of time. There is therefore room in Hummel and Holyoak’s theory underlying LISA for the kind of representational change propose here: change that’s produced by forging the analogy in the first place.

In the service of learning, many kinds of representational change are explored by analogy researchers. Most of these are longer term changes that take place over more extended amounts of time. DORA (Doumas et al. 2008) is one excellent example, for it can learn novel relational concepts from examples. Since DORA is a dynamic, connectionist model, it routinely does a kind of representational change: it builds up its active representations from what’s in memory, and what’s built is not an exact duplicate of what’s in memory. Furthermore, DORA, again in the service of learning, builds abstractions: it can derive the stand-alone predicate *big* from examples of a big bear and a big house (see, Doumas et al. 2008, Fig. 6). Another example of a representational-change learner is SEQL (Skorstad et al. 1988; Kuehne et al. 2000). SEQL (and its associated generalization and exemplar learner, GEL, Kuehne et al. 2000) preserves only that information that lies in the intersection of two concepts or representations. This is clearly a kind of abstraction. The kind studied here, rapid abstraction, is similar to the kind in SEQL but differs from it primarily in that the intersection of two representations is not the only kind of change employed (this will be explained in “Rapid abstraction in STRANG”). Furthermore, as mentioned in note 3, the mechanism for rapid abstraction works as part of the analogy-making process and could be inserted into various analogy engines, such as MAC/FAC; SEQL on the other hand is part of a learner which does progressive abstraction, which, again, occurs over extended time, whereas rapid abstraction occurs early and quickly. Of course, SEQL could be altered and made a part of, e.g., MAC/FAC, but it would still use progressive abstraction and so function differently from rapid abstraction.

Finally, abstraction as a form of representational change is nearly ubiquitous throughout cognition (e.g., learning, schema abstraction, many forms of concept formation, induction, and categorization), and it is well studied in cognitive science. It is also very likely not just



one kind of representational change (see, e.g., Barsalou 2003).<sup>5</sup>

Why rapid abstraction is needed for analogy:  
pre-identical structures and the low probability  
argument

There is relatively wide agreement that analogy includes *retrieval* of the base domain (e.g., comet orbits) and *mapping* it to the target (e.g., alpha particle paths). (Some models sharply distinguish between these two, e.g., MAC/FAC; others do not, e.g., LISA.) Mapping is a process of aligning object representations and relational structures from one mental representation to another, while usually jettisoning many of the property representations (this is itself a kind of abstraction). Very simply and abstractly put, if  $R_1(F(a) G(b))$  and  $R_2(H(c) J(d))$  are (highly simplified) mental representations with objects  $a, b, c, d$ , having, respectively, properties  $F, G, H$ , and  $J$ , and being in relations  $R_1$  and  $R_2$ , then an analogy between the two maps  $R_1$  to  $R_2$ ,  $a$  to  $c$  and  $b$  to  $d$  but jettisons at least some of the properties  $F, G, H$ , and  $J$ . Hence, analogy (analogical mapping) is essentially a process of finding *functional counterparts* between concepts (see, e.g., Gentner 1983, 1989; Hummel and Holyoak 1997).

As we have seen, two crucial assumptions of many models (but not all) which otherwise have very little in

common are that, *at the time of retrieval*, the structural relations  $R_1$  and  $R_2$  are *identical* and *exist ahead of time*.<sup>6</sup> The analogy, therefore, is in fact between  $R_1(F(a) G(b))$  and  $R_1(H(c) J(d))$ .<sup>7</sup> Both representations already exist in memory in that form (one in working memory, the other in long-term memory); this is needed to make an analogy. Other models do not require identical structures at the time of retrieval, but nevertheless, require robust semantic similarity. LISA (Hummel and Holyoak 1997) is one such model. In fact, LISA does not draw a sharp distinction between retrieval and mapping. Still, even LISA's analogies depend on strong semantic similarities. For example, the 1997 version of LISA finds analogies between, e.g., "John loves Mary" and "Bill likes Susan" (see Fig. 4 in the 1997 version). (Note that LISA requires semantic similarity only between the roles or relational predicates; the objects involved need have no semantic similarity.) The 2003 version extends LISA to include inference and generalization but still requires high semantical similarity.

The question we have to ask now is: What is the probability that two complex relational representations, from completely different, semantically distant domains, have identical or strongly similar pre-existing structures and semantics? Prima facie, the probability is too low to explain the abundance of spontaneous analogical

<sup>5</sup> There are other important models of analogy which hypothesize various kinds of representational change, some of which are interpretable as structural changes, others are interpretable as abstractions similar to the kind considered here. AMBR2 is one such model [although Kokinov and Petrov seem to downplay within their theory the importance of structure in analogy; see Kokinov and Petrov (2001, especially, tables 3.1 and 3.4)]. Also, AMBR2 seems to assume a pre-existing semantic similarity. This might be theoretically fine, depending on the details of "semantic similarity." In any case, AMBR2 can map together quite non-similar predicates, but this is not due so much to representational changes in the predicates, but rather to special interpretations of the "semantics" of the relevant predicates. Other representational change analogy theories include COPYCAT (Mitchell 1993) and TABLETOP (French 1995). In these two, the representational changes occur during the making of the analogy and are roughly similar to the kind proposed here—rapid abstraction (COPYCAT has been criticized as not being an adequate model of analogy for other reasons; see Forbus et al. 1998). Their criticisms also apply to TABLETOP, for COPYCAT and TABLETOP are deeply similar, even analogous). Indurkha has an interesting analogy model using representational change (see his 1992, 1997, 1998). Finally, some robust and highly suggestive computer models of analogy and representational change are based on psychological data collected in experiments. The performance of these models has also been favorably compared with human performance. In many ways, these programs and their ties to psychological experiments represent one of cognitive science's real success stories. See ACME and its associated model of memory access, ARCS (Holyoak and Thagard 1989; Thagard et al. 1990), IAM (Keane and Brayshaw 1988; Keane et al. 1994); Phineas (Falkenhainer 1990a, b), MACFAC (Forbus et al. 1995), LISA (Hummel and Holyoak 1997, 2003) and DORA (Doumas et al. 2008).

<sup>6</sup> For example, compare Structure-Mapping Theory (e.g., Gentner 1983, 1989) to the developmental, connectionist relational priming model of Leech et al. (2008). The former well-known theory and accompanying computer model use classical propositional-style representations. Analogous representations are said to be in *structural alignment*. This alignment is the basis for the analogical mapping, and it assumes that the aligned elements already exist. Note that the three key constraints in SMT governing structural alignment—*parallel connectivity*, *one-to-one mapping*, and *systematicity*—are based on the assumption of pre-existing identical relational structures: the three constraints govern the best *match* between such representations. Parallel connectivity ensures that when two elements are placed in correspondence, the arguments of those elements are also placed in correspondence. The one-to-one mapping principle states that each element in a base domain can be mapped to at most one element in the corresponding target domain. The systematicity constraint requires that, all else being equal, correspondences between *systems* of elements in the domains are preferred to matches between isolated elements. The mapping process is local-to-global. At the beginning of the comparison, individual entities, attributes, and relations are matched, and then the constraints are utilized to determine more global systems of relations. In contrast, Leech et al.'s model of analogy uses a recurrent connectionist net that relies on relational priming and pattern completion. It's main focus is on developmental and learning issues. Relations are represented as transformations between states and are stored in the weights and hidden layer. In spite of these differences between the two models, Leech et al.'s model explicitly uses pre-existing, identical relations to fund the analogies their network makes. Their analogies are just as discovered (and not created) as those in SMT (see Dietrich 2008).

<sup>7</sup> Turney (2008), provides an enlightening discussion of the problems of pre-existing and hand-coded representations. His Latent Relational Mapping Engine is his proposed solution.

reminders. Two related considerations are operative here. First, spontaneous, insightful analogies almost always use analogs that are very different semantically. Dog tongues and human hands differ a great deal (at least they do outside of the context of the analogy given above). Atoms, solar systems, alpha particles, and comets differ even more (again, they do so if we ignore the analogy that relates them). Second, this semantic difference manifests itself both in representational structure and objects. A high degree of semantic similarity of the kind needed for many models of analogy is, therefore, unlikely to be available in natural settings. The only way out of this, ignoring the semantics and focusing only on the syntax, lands us in the opposite situation. For example, if mere identity of predicate arity is sufficient, then, e.g., all 2-place predicates are going to be analogous, leading to a troubling surfeit of analogies which would render analogies all but useless, and which, in any case, we do not find in actual everyday human thinking. However, ignoring the semantics is not really an option for many other reasons as well.

Recall that Rutherford's alpha particles were by no means moving around the nucleus the way comets move around the sun, and Rutherford knew this. The particles were deflected by the constituents of the gold foil. And Rutherford knew this fact, too: this was the basis of his comment about the fifteen-inch shell and the tissue paper (eventually, Rutherford and others would conclude that the alpha particles were strongly interacting with the gold atoms' electrostatic charge). Pretty obviously, the relevant relations, "orbits" and "deflects" are not identical, nor are they very similar, lying as they do in such vastly different domains as comets and alpha particles.<sup>8</sup>

This observation—that the probability of antecedently identical structures in semantically different domains is too low to account for the observed quantity of spontaneous analogical reminders—forms the basis of my claim that the relevant representations have to change in order to forge an analogy. I call this argument the *Low Probability Argument* (see Dietrich 2000, where I handle several objections to the argument; see Dietrich et al. 2003, where the argument is bolstered by considering the analogy between heat flow and water flow, and by considering analogies in Kepler's development of his theory of planetary motion). Structure mapping theory as well as many other analogy theories are committed to the view that almost all of the structural representational changes occur *because* of the analogy. The Low Probability Argument suggests that this is very unlikely: some representational changes are crucial to *making* the analogy.

Consider Rutherford's analogy again. He's thinking about the surprisingly deflected alpha particles and is reminded of comets' sharp turns around the sun. The Low Probability Argument claims that it is very unlikely that Rutherford's concepts of deflected particles and comet paths shared much structure or structural information (semantics) before these concepts or representations became analogues. But if they share a lot of structure and content after they become analogues, where did the shared information come from? Now it is time to look at the proposed representation change—rapid abstraction—in more detail.

### Rapid abstraction in STRANG

STRANG is a computer program that implements the sort of representational change being considered here. The idea behind STRANG is *abstracting via representational interaction*. STRANG implements a part of an algebra (in the universal algebra sense; see, e.g., Lipson 1981). STRANG's domain is the letter string domain; the strings conform to a grammar and are almost bereft of semantics. STRANG builds structured representations by parsing input strings according to the fixed grammar (though the grammar is fixed, the eventual abstractions do not have to adhere to it, as we shall see, below). Call the structured representation built from the input string  $T$  (for "target"), and call the structured representation in long-term memory  $B$  (for "base"). In STRANG,  $T$  becomes an algebraic operator applied to  $B$ . So, we have  $T(B) = A^*$ , the abstraction or alteration of  $B$  relative to  $T$ .<sup>9</sup>  $A^*$  is  $B$  with all of its detail relative to  $T$  packed away (as discussed, this is similar to SQL, but the changes are more than just intersections, as explained in this section). *Packing* is the formal algebraic operation implemented in STRANG (Dietrich, et al. 2003).

The easiest way to explain STRANG's abstracting procedure is to offer some examples. The details of the parsings and representations will be skipped (these can be found in Dietrich et al. 2003, and Winkley 2006). Just the input and output will be described.

Given  $T = abab$  and  $B = efef$ , STRANG represents their algebraic combination (via packing) as a kind of tree, called a packing tree, whose root node is  $\alpha\alpha$  (note, these packing trees are *not* parse trees). This means that STRANG abstractly regards both strings as something,  $\alpha$ , that repeats (this is their "essence," as it were). The string  $\alpha\alpha$  represents the original strings maximally packed, while still preserving what's important about them. Both  $\alpha$ 's unpack into  $2SQ$ , which means: "a sequence two letters long." Hence, at a lower level, STRANG regards both strings as repeating sequences two letters long. The  $2SQ$ 's unpack (in this case) into (ab) and

<sup>8</sup> The fact that "orbits" and "deflects" are not identical would not be a problem for models such as LISA, but the fact that they are semantically so dissimilar would be.

<sup>9</sup> COPYCAT (Mitchell 1993) does something roughly like this.

(ef), respectively (with their string membership kept distinct and intact). The result, which represents the various packings and unpackings, is the tree:

$$(\alpha\alpha(2SQ(a)(2SQ(e))))^{10}$$

(Since the length is known, only the beginning of the sequence is needed in the representation of the packing tree. For example, the first  $2SQ$  unpacks into the sequence (ab).) In short-hand, the strings are represented as ((ab)(ab)) and ((ef)(ef)).

Given  $T = \text{efgefg}$  and  $B = \text{abab}$ , STRANG again produces  $\alpha\alpha$  (with the same meaning as before). So again, STRANG regards both strings as essentially repetitions of something. The left-most  $\alpha$  unpacks into  $3SQ$  and the right-most to  $2SQ$ , with  $2SQ$  unpacking into (ab) and  $3SQ$  unpacking into (efg). The resulting tree is:

$$(\alpha\alpha(2SQ(a)(3SQ(e))), \text{ i.e., } ((\text{ab})(\text{ab})) \text{ and } ((\text{efg})(\text{efg})).$$

More interestingly, given  $T = \text{ababccc}$  and  $B = \text{mnopqrhijhijhij}$ , STRANG produces a novel abstraction, resulting in the trees:

$$(((\text{ab})(\text{ab}))(\text{ccc})) \text{ and } (((\text{mno})(\text{pqr}))(\text{hij})(\text{hij})(\text{hij})))$$

(the technical details are in Dietrich et al. 2003). STRANG represents both strings as “two same-length sequences followed by a 3-item repeating string.” There are two interesting and novel things about this representation, both of which violate STRANG’s grammar-based restrictions. First, STRANG’s grammar does not allow for parsing the string (mnopqr) into (mno)(pqr); straight letter sequences cannot, according to the grammar rules, be broken up into parts. Yet this was done. This first part of the parse was produced under the influence of the substring (abab). The abstraction here is that the strings (abab) and (mnopqr) are viewed by STRANG as basically the same: they are “repeating” letter sequences; the actual letters do not matter, though, nor do the lengths of sequences. It does not even matter if the lengths are the same across the two differing sequences (in this case, two letters versus three letters). All that matters is that there are sequences of given lengths and it is the *length* which repeats. Specifically, if *Length* is a function which returns the length of a sequence, then the abstraction at work here is this: given sequences  $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ ,  $(\sigma_1\sigma_2)$  is analogous (in this particular case) to  $(\sigma_3\sigma_4)$  provided that  $\text{Length}(\sigma_1) = \text{Length}(\sigma_2)$  and  $\text{Length}(\sigma_3) = \text{Length}(\sigma_4)$ . So, ((ab)(ab)) and ((mno)(pqr)) are analogous because they are both sequences of the same length (2 and 3, respectively). (It might be objected this is quite a thin reason for making the two analogous. By itself, the *Length* analogy would be thin, and STRANG would not

make it. But in the context of the rest of the string (see the next paragraph), constructing the *Length* analogy is strongly motivated. It is this notion of *context* that partly supplies STRANG with what little semantics it has.)

The second grammar violation in STRANG is that repeating strings consist of only one letter; (ccc) is such a legal string but ((hij)(hij)(hij)) is not. Nevertheless, STRANG “sees” these strings also as basically the same. In a sense, STRANG invented a new relational class, interpreted (by humans) as “a 3-item string.” Hence, after the construction of the analogical abstraction, STRANG “thought” of (ccc) and (hijhijhij) as abstractly the same: strings made up of 3 identical parts. It abstracted away the fact that the parts were not similar at all across the two strings. In terms of mapping, each “c” was mapped to an “hij”.<sup>11</sup>

These packings and alternations produced by STRANG are examples of rapid abstraction—the kind of representational change that this paper argues is needed to more fully understand analogy.

One objection to rapid abstraction is that it is too expensive, time-wise, to be generally useful, so it cannot be involved in daily analogical reminders across large distances. However, the kinds of changes STRANG produces might be got almost for free using another important fact about human cognition.

### Abstraction and psychological distance

In series of important papers, Liberman and Trope have proposed and explored an interesting theory about how humans transcend the here and now (for a summary see, Liberman and Trope 2008; for a recent extension of their theory, see Liberman and Förster 2009). The theory spans quite different time scales, from human evolution, to human history, to child development to human categorization. This last is the part of their theory that is relevant here. The core of their theory is this: “different dimensions of psychological distance—[e.g.,] spatial, temporal, social, and hypotheticality—correspond to different ways objects and events can be removed from the self, and the farther removed objects are

<sup>10</sup> All trees are binary and are represented as lists thusly: [root (left-subtree) (right-subtree)]. Also, the nomenclature used here is simplified from that used in Dietrich et al. (2003).

<sup>11</sup> In STRANG, mapping is constructing a link between the mapped items. This is implemented usually as coupling in a list. It is important to emphasize here that STRANG’s packing trees can violate the rules of its grammar. Of course, STRANG can do this because of the algebraic operation it implements, namely, *packing*:  $T(B) = A^*$ . Packing implements implicit rules that can override the rules of the grammar. These implicit rules are derived from each operator string,  $T$ , as it is applied to its base string,  $B$ . When the grammar is overridden, STRANG usually marks this explicitly by a specific new node type in its packing tree. There is only one such new node type, but it can be employed in a wide variety of cases. This new node type represents that STRANG has a new parsing category for which it does not have a name and which is not integrated into the grammar.

represented at a higher, more abstract level,” (2008, p. 1202). So basically, the further away something is from the self’s here and now, the more abstractly it is represented (2008). (The self’s here and now (or just the here and now for the person involved) is always the *origin* from which the relevant distances are computed.) Put succinctly: psychological distance and abstraction are linked covariantly. Somewhat more strongly, psychological distance *induces* abstraction. This seems intuitively correct: from a distance we see the forest, up close we see the trees.<sup>12</sup>

According to Liberman and Trope, any object or event can be represented relatively concretely or at higher levels of abstraction. Concrete representations are less structured, more contextualized, and contain more information in the form of incidental features. Higher level abstract representations are schematic, decontextualized, and tend to represent the gist of an object or event by focusing on core features and omitting incidental information. As the psychological distance of what is being represented increases (relative to the self at a specific here and now), so too does the represented abstraction of that thing. Moreover, many of the psychological distances are closely related. For example, something represented as far away in time tends also to be represented as far away in space (2008).

To apply Liberman’s and Trope’s theory to insightful analogical retrieval we need two things: (1) a point of origin, i.e., a here and now relative to which memory items are closer or farther away, and (2) a relevant psychological distance. We can extend Liberman’s and Trope’s theory by adding in the dimension of *semantic distance* between concepts. For example, the semantic distance between atoms and solar systems is larger than the distance between atoms and molecules. So, the relevant psychological distance here is semantic distance. This gives us (2).<sup>13</sup> The origin, or the here and now aspect, is fixed by the concept that is currently the main focus of attention. So, for example, if someone is currently thinking about atoms, that is the origin. Relative to it, conceptual knowledge about solar systems is far away, semantically speaking. This gives us (1).

<sup>12</sup> A detailed explanation for *why* psychological distance covaries with and induces abstraction is missing from Liberman’s and Trope’s theory (but see their 2008, p. 1205, and also Liberman and Förster 2009, p. 1337, for some interesting speculation). Adding such an explanation to their theory would be a major advance. Also missing is an actual mechanism for producing the abstraction. Adding this would also be a major advance.

<sup>13</sup> I’m leaving the notion of semantic distance intuitively defined because a full definition would require a theory of mental semantics or mental content (including intentionality). In Liberman’s and Trope’s theory, their notion of psychological distance resolves into many kinds of specific distances from self (origin), including distances of time, space, probability of occurring, social relations, etc. Each different kind of distance has a different metric; yet, all the different kinds of distances are related. See their 2008.

Now, we get the needed rapid abstractions basically for free. To see this, consider Rutherford again, just before he had his analogy. Atoms are here and now for him because they are currently his focus of attention. So, according to our extension of Liberman’s and Trope’s theory, any concept from a domain of concepts semantically far away will be, when retrieved, already abstractly represented simply because semantic distance induces the relevant abstraction (which is what all the other psychological distances Liberman and Trope consider do).

Returning to the last STRANG example from above, suppose that  $T = ababccc$  is in short-term memory and is the focus of attention, and suppose that  $B = mnopqrhij-hijhij$  is a concept from long-term memory. STRANG’s grammar implements a kind of *syntactic distance* that can be interpreted as similar to semantic distance in human analogies:  $B$  is “semantically” quite distant from  $T$  due to the deep differences in their parse trees, which can be measured by the fact that  $T(B)$ , in this case, results in violations of STRANG’S grammar, as discussed earlier. (Of course, as also mentioned, the semantics of STRANG’S grammar is quite thin, and derives from its syntax. This is not as radical a move as it might seem, for syntactic differences can *be* semantic differences, see Rapaport 1988, 1995, 1999.) Upon retrieval of  $B$  given  $T$ , a Liberman and Trope version of STRANG could produce the novel abstraction discussed earlier, i.e.,  $((mno)(pqr))((hij)(hij)(hij))$ , merely in virtue of the semantic distance between “ababccc” and “mnopqrhijhijhij.” On the proposal being made here, the phenomenon of rapid abstraction turns out to be a *consequence* of semantic distance, and as such is got relatively cheaply, computationally speaking.<sup>14</sup>

<sup>14</sup> We know from Liberman and Trope’s research that retrieval across large psychological distances either quickly *produces* or *merely retrieves* abstractions. It seems unlikely that the relevant abstractions are already made and are just lying around (for one thing, each concept would have to be abstracted relative to all other concepts in a given mind; that’s a lot of abstractions, to put it mildly). So, we can conclude that large distance retrieval produces the abstractions. How? As mentioned above, Liberman and Trope do not speculate on the actual mechanism by which abstractions are created by psychological distance; they are concerned primarily to argue that such abstraction occurs and is due to psychological distance. Perhaps the STRANG algorithm could be the needed mechanism. Briefly, given some cognitive agent and any two of its mental representations,  $R$  and  $S$ , these two differ as a function of their semantics. Hence, representations that differ a lot are going to be semantically further apart than representations that differ only a little. Ergo, if  $R$  is an operator on  $S$  in the way described in this paper, the resultant representation,  $U$ ,  $R(S) = U$ , will be more and more abstract as the semantic distance between  $R$  and  $S$  increases (compare the packing tree representations of  $abab$  and  $efef$  versus  $ababccc$  and  $mnopqrhijhijhij$ ). The problem lies in making the STRANG algorithm fast enough in an architecture such as the brain. This might be easy or hard. At this point, we do not know. In fact, we do not in detail know how the brain implements any high-level cognitive processing, analogy included.



Here is what we have so far: (1) the introduction of a kind of representational change called rapid abstraction and an argument that it is required for a certain kind of analogy—insightful, spontaneous analogical reminders occurring over large semantical distances; (2) a general method by which such abstractions can be cheaply produced: the Liberman–Trope theory. But one major goal of this project was to unify categorization and analogy. So now we need a way to view categorization (at least some categorizations) in such a way that the rapid abstraction can participate in a kind of categorization. “[Construing and categorization](#)”, begins this, and “[Construal-based analogy: unifying analogy and categorization via construing](#)” completes it.

### Construing and categorization

The goal of *construing theory* is to explain how cognitive agents construct meaningful *understandings* of objects in their environments even though they begin with meaningless, semantic-free perceptual representations of those objects (i.e., perceptual input). In other words, the goal is to go from semantic-free representations to semantical, and hence *understood*, representations. Semantical representations are how we know our environment. First, I will describe how construing works. Then I will connect it to categorization (for more detail, see Kurtz and Dietrich 2006).

In essence, the goal of the construal process is to merge generic, categorical information from long-term memory, which is very definitely semantical, in a strict sense, with semantic-free (but potentially semantically crucial) perceptual information from a cognitive system’s environment. So, the output of the construing process is a semantic understanding of an encountered object of some sort, i.e., the object, say, a dog, is understood *as* a dog.

The sense in which concepts are (or categorical information, in general, is) semantical is that such information can be used to make *inferences* of all kinds (“if X is a dog, X might bite, X might chase cats, X is a mammal, X is related to wolves, X’s sense of smell is very robust,” etc.) In short, the semantical aspect of categorical information provides *knowledge*. Importantly, there is another key aspect to semantics beside inferential potential, and that is *informational contact* with an agent’s environment. Concepts do not readily provide such informational contact, perceptual information does, however.

There are four main steps in the construing process. Consider the case of encountering a dog. The first step involves the minimal and non-controversial assumption that the input sensory information is converted into an *initial visual description*: i.e., a detailed and largely veridical, *but*

*meaning-free* representation—such as a sketch of parsed collections of edges or a structural description based on a vocabulary of (semantically *uninterpreted*) geometric components and spatial relations. It is important to see that when this step is completed, a visual description (representation) exists, but nothing meaningful, nothing semantical has been constructed yet. How could it? Information (say, in the form of light) enters the eyes and the early visual system constructs a perceptual representation. Unless the world came with pre-packaged semantic information (it does not), there is no way for this step to contain any semantics, and meaning. Of course, it does not follow, and it is not being asserted, that the information at this level is *not useful* for constructing a semantical interpretation of an object (i.e., *what* that object is). The perceptual information at this level is crucial to constructing a full, semantical interpretation.

Second, the initial visual description is evaluated to produce an object recognition decision such as: “looks like a dog.” This decision, which is dubbed the *summary description* provides an *uninterpreted label* pointing to a category. The summary description is also decidedly *not* a semantic interpretation of the visual description because the system supplying the description, e.g., “looks like a dog,” does not know what a dog is; it knows only that the current perceived object looks like one of its output classes. (Using as an example of a summary description the proposition “looks like a dog” might be confusing for, of course, the proposition has meaning, but this is an artifact of giving the example in English.) It is useful to think of this part of the construing process in terms of standard objection recognition software. Such software systems do not begin by recognizing features like ears; this would be getting the semantical cart before the syntactical horse. They begin by recognizing purely syntactical features like edges or holes or geometrical shapes (e.g., geons in recognition-by-components; Biederman 1987). This lack of semantics means that the describing system supplies no *knowledge* about what the thing being perceived is and cannot produce inferences or any knowledge-driven interpretation of the perceived object. The label “looks like a dog,” for example, does not entail “not a cat,” nor “is a mammal,” nor “is probably friendly,” etc. The label is semantically inert.

Third, the summary description derived from perceptual appearance is used to access long-term memory and retrieve a generic category such as *dog*. This retrieval does not function like a data-base query in which each stored category is compared to a probe. Instead, the summary description acts as a direct pointer into long-term memory. As such, category access is a simple, rapid look-up procedure. I assume that the generic category information is represented as a *schema* that is that the generic information

is represented by slots, which are either open or filled with default fillers. For example, the *dog* schema has a slot for fur color, which perhaps defaults to brown (for the particular individual possessing that *dog* schema). The default fillers function as suggestions. They are easily altered as the details of the actual scene in the agent's environment warrants—this takes us to step four, the core of the construing process.

The fourth and final step of construal is to take the accessed generic category—which holds semantic (i.e., inferential) knowledge of dogs in general—and map the initial visual description into it (so the initial visual description is used twice: once to make the summary description and once to flesh out the retrieved category with specific real-time information, thus being *construed* by the retrieved category). Specifically, the slots in the schema are filled in with information from the visual description. In this way, the perceptual description of the stimulus is *interpreted* relative to the generic category information. The interpreted visual description is the *construal*. The information in the construal takes the form of (e.g.): “a dog that has a malformed tail,” “a dog with white fur,” “a dog that is very large and barking,” and so forth. In this way, the dog in the environment is understood *as* a dog; it is seen *as* a dog with certain dog-relevant properties (is barking, has white fur, is large, etc.)<sup>15</sup> The construal is thus the *meaningful* categorization of the perceived object: an organized set of representational elements that constitutes a semantic interpretation of the visual evidence based on a stored category.

For purposes of illustration, a concept for *dog* might include slots and defaults like this:

DOG:  
 has-a-tail [medium length]  
 has-fur [brown]  
 has-size [medium]  
 makes-sound [bark]  
 has-behavior [friendly]  
 has-legs [4]  
 is-a [animal]

When an actual dog is encountered, it might be consistent with this concept in terms of its legs and tail, but it may also happen to be large, white, and growling.

<sup>15</sup> The phrase “see as” is meant to flag the distinction between seeing *that* something, X, has property F, versus seeing X *as* F. This distinction has a long tradition in philosophy of mind and is notoriously hard to make clear. Nevertheless, it can be useful. It is roughly the distinction between just seeing something and interpreting it a certain way. For example, one might just see two faces where someone else sees a family resemblance between the two faces. The latter case is a case of seeing-as. This is arguably a species of what Jerome Bruner called “going beyond the information given,”—see Bruner (1957). For a study of the relationship between Bruner's notion and seeing-as, see Osborne and Lee (1997).

Construing proceeds thusly. (1) The initial visual description of the stimulus is constructed; it consists of perceptual information derived from the dog in the environment and the initial description does not convey meaning. (2) The object recognition system generates a purely visual identification—the summary description—that (3) provides access to the *dog* concept as a candidate category; *dog* is then retrieved for use in the construal of the stimulus. (4) The initial visual description is then mapped onto the generic concept. In this way, the semantic slots will be assigned values *from the perceptual data*; so a semantic encoding of the stimulus is constructed by *integrating* generic knowledge and available perceptual data.

To summarize, construal-based categorization consists of the following steps:

- Step 1. Create a non-semantic, initial stimulus representation via low-level visual information processing.
- Step 2. Generate a summary description based on a non-semantic process of high-level visual object recognition.
- Step 3. Retrieve the category designated by the summary description.
- Step 4. Integrate, via mapping, the initial visual stimulus description and the generic category knowledge to produce a construal.<sup>16</sup>

Finally, is categorization usefully explained as construing, at least some of the time? Yes. And for basically one reason: *semantics*. Categorization is usually theorized about in such a way that the crucial semantical questions are elided, leaving a glaring hole in the theory: how does the perceptual-categorizing system go from *semanticsless* input to semantical and *fully meaningful* mental representations? We know this semanticsless-to-semantical transition occurs. We know categorization is deeply involved. It is not much of a stretch to conclude that a theory of categorization should explain this transition. But no well known, established theory does (see Kurtz and Dietrich 2006, where this is argued in detail).

There are two *non-construing* ways out of this problem, neither of which work. First, a theory of categorization could *begin* with semantically interpreted input, i.e., it could deny that there is a *semanticsless* to semantical transition. This way out, however, is hopelessly question-begging. Any theory like this would be assuming an understanding of the stimulus in order to categorize it. But explaining the emergence of such understanding was the original goal of theorizing about categorization in the first place. So assuming semantically interpreted input would be assuming what categorization theory is supposed to explain.

<sup>16</sup> The construing model is not only theoretical cognitive science and philosophy of mind. Part of it has been implemented and tested. See Kurtz (2007). Other experimental tests are discussed in Kurtz (2005).

It might be objected that there is a way for the above to work. The *features* associated with the input could be identified and then the category with those features could be retrieved. If one is looking at a dog, the first step could be to identify ears, fur, legs, and tail, and then retrieve the concept having those features. But figuring out that something is an ear, a leg, or a tail *is* categorizing it. So, on this variant of “assume interpreted input,” categorizing something would require first categorizing its parts. This leads to an infinite regress.

There are theories which have the consequence that information of a specific sort is already semantical, or already contains semantic content. Dretske’s information-based knowledge is one such theory (Dretske 1981). But such theories still require active mental processes to recover the semantical information. All such theories do show that the semantical information is already there to be recovered; the mind, on such theories, cannot just passively receive the input and thereby have semantically interpreted information. The mind has to *do* something.

The second way out of the problem is to leave the perceptual input uninterpreted and then compare it to every category in long-term memory to see what it is, categorically speaking. This way, too, is hopeless. The time-complexity of such a comparison process, even if done in parallel would render categorization too slow to be of use (see Tsotsos 1990, 1995; in his 1990, Tsotsos also argues that issues of time-complexity render useless any purely bottom-up approaches to visual identification of the kind the first way and second way are committed to). Another problem with this approach is that it does not use the perceptual information in deriving an understanding of a perceived object. The information is just used for retrieval.

If explaining semantics—how our percepts acquire content and come to be understood—is a central goal of theorizing about categorization, then the semantics cannot be assumed up front without begging all the important questions, and the semantics cannot be searched for in the time limits required. This seems to leave alternatives where the semantical information is *constructed*. Construing theory is just such a theory.<sup>17</sup>

<sup>17</sup> Another argument for construing theory is that it nicely fits with many current psychological and philosophical approaches of semantics in cognitive systems. These approaches divide semantics into two dimensions: an internal dimension and an external one. The internal one is based on the implicative or inferential power of mental representations. The external one is based on informational contact with the world. Both dimensions are needed for a full semantical connection to one’s environment (See Dietrich 2006, and Markman and Dietrich 2000, 1998, where this is argued in detail). Construing uses both dimensions and so can be said to be fully semantical. For a given construal, the internal dimension is realized by the generic concept’s inferential potential. The external dimension is realized by mapping the initial visual description (derived from perceptual input) onto the generic concept.

## Construal-based analogy: unifying analogy and categorization via construing

What is required now is a way to marry rapid abstraction with construing. This will give us our unification of the sort of analogy-making theorized about here and categorization, at least those categorizations that are usefully described by construing theory.

In standard construing, the summary description retrieves the relevant category. But this would not let us use the Liberman–Trope theory, since the summary description cannot function as our point of origin, because, as discussed in “[Abstraction and psychological distance](#)”, it has no semantics—it is just a pointer into long-term memory (hence, no psychological distance is involved). To use Liberman–Trope theory, we need a semantical point of origin. In “[Abstraction and psychological distance](#)”, I suggested that the concept currently the focus of attention could be the point of origin. It is natural, therefore, to take *a currently active construal* as the point of origin. For example, in Rutherford’s case, a plausible currently active construal is his thought about the aberrant paths of the alpha particles. Calling this a “currently active construal” just means that it is the result of a standard construal-based categorization of the type considered in “[Construing and categorization](#)”: Uninterpreted perceptual input about the aberrant paths of the alpha particles was construed by Rutherford’s category of alpha particles and his knowledge of the how the experiment was set up.

If this currently active construal, in turn, is used as a point of origin to retrieve a semantically distant category (e.g., comets and their paths) from long-term memory, that category will be automatically and rapidly abstracted because of the Liberman–Trope idea: “distance equals abstraction.” (Note 14 is relevant here.)

Now, the semantically distant category, “comets and their paths”, can participate in *another* construal, this time between it and the point of origin, in this case, the aberrant paths of the alpha particles. The claim here is that the product of a standard categorization construal can be used in *another* construal (sort of a “meta” construal). It is important to note here that the analogy is not really complete until the second construal is finished. So, this type of analogy produces a *product*; it is not just a mapping between two mental representations in memory. The product is the second (meta)construal. (STRANG only models the rapid abstraction part.) The result is a spontaneous analogical reminding occurring over a large semantical distance that produces an insight.

So, if spontaneous analogical remindings occurring over large semantical distances are correctly viewed as

involving (meta)construings, and if construing theory correctly describes certain categorizations, we have our result. To the extent that construing is a form of categorization (and, as argued in “[Construing and categorization](#)”, this extent is at least not zero due to needing to explain the emergence of semantical understandings), we have the unification of a certain class of analogies and a certain class of categorization.<sup>18</sup>

### A simple example

We are now ready to look at the comet-alpha-particle analogy in some detail. We are not here interested in the entire analogy-making, analogical reasoning process whereby Rutherford concluded that most of the mass of an

<sup>18</sup> An objection to the *rapid abstraction-construing theory of analogy* is that it seems to imply that far analogies are easier to make than near analogies. It has been suggested that work such as Gick and Holyoak’s (1980) seems to tell against this. The quick answer to this is, Yes, the greater the psychological distance, the lower the analogical retrieval time, *other things being equal*, and *assuming that abstractions are handled faster*; but, No, Gick and Holyoak’s results are not a problem for this. Here is a longer answer, in three steps. First, field data indicate that spontaneous analogical reminders are rather common (Holyoak himself, together with Thagard, makes this point in their 1995). But they are rare in the psychology lab. Dunbar calls this the *analogical paradox* (2001). As I said at the beginning of “[Semantically distant analogies and rapid abstraction](#)”, I follow Dunbar (2001) in the view that natural spontaneous analogical reminders are quite common. Dunbar explains away this paradox by pointing out that lab settings for analogy production are actually quite a bit different from natural settings in several ways. One of those ways is that in natural settings, people *generate* analogies, rather than merely chose them, as they often do in lab settings. The research reported in this paper is geared toward analogical generation. Second, the research reported here does seem to suggest that given four concepts, C1, C2, C3, C4, if the psychological distance between C1 and C2 is greater than the psychological distance between C3 and C4, then C1(C2) [the rapid abstraction operation] will happen faster than C3(C4). But this is just a general, loose statement. It is not clear that rapid abstraction-construing theory predicts this because (1) factors like the difference in complexity of the operations C1(C2) and C3(C4) might swamp any distance effect [e.g., C1(C2) might be more complex on average or in certain cases than C3(C4)], and (2) construing times will not affect C1(C2), but will affect the time for final generation of the analogical insight, and rapid abstraction and construing are very tightly coupled; it is not clear how to separate them, experimentally. Given (1) and (2), it is not obvious how to set up an experiment to test this “distance-is-quickier” hypothesis, and it is not clear what a negative result would mean. Third, Gick and Holyoak never explored the role of *relative* distance between concepts in their famous experiments. Their tests were concerned only with some sort of “absolute” or fixed distance, namely that presupposed in Duncker’s Radiation Problem. Also, Gick and Holyoak only found that, *in the lab*, a certain distant analogy was hard to make. Besides the problems Dunbar raises in his 2001 paper, in their experiments, Gick and Holyoak wanted their subjects to find a *specific* analogy. They did not experiment on open-ended, spontaneous analogical reminders.

atom is concentrated in a very small, positively charged region (called the nucleus). We are only interested in that very first insight: alpha particles hitting gold foil behave like comets. This insight was the trigger for the longer, more involved analogical reasoning that lead eventually to a better model of the atom.

Return to Rutherford in his lab.<sup>19</sup> He has witnessed the highly deflected paths of the alpha particles. How was he analogically reminded of comets? The following steps characterize, I claim, Rutherford’s analogical reminding, and they occurred rapidly. (No claim of historical accuracy is being made for the steps; yet, as history, they are not completely implausible. Also, and again, I am theorizing about rapid abstraction and construing as used in analogy, not complex, long-lasting knowledge change over time, like learning and reasoning.)

Step 1—*Initial surprise*. Rutherford is surprised that some of the alpha particles were highly deflected by the gold foil; basically, they bounced off of the foil. This should not have happened. At this step, Rutherford understands atoms only on the Thomson model. This model has the atom’s electrons embedded in a large, positively charge matrix. The image was of a positively charged soup-like or pudding-like mass with negatively charged electrons stuck in it (like plums in a plum-pudding, hence the name). Thomson atoms are not dense enough, basically, to explain Rutherford’s experimental results. Let us suppose that the following schemas were active (note: a schema’s name is considered part of the schema):

---

#### Bounced-off

**type:** relation

**default relata types:** small

projectile, sufficiently resistant surface

**small projectile:** *alpha particles*

**resistant surface:** *gold foil*

#### Gold foil

**type:** object

**composed of:** *gold atoms*

**salient property:** thin

**property “thin” specification:**

400 atoms thick

---

#### Alpha particles

**type:** object

**salient property:** fast, too fast to be sharply deflected by diffuse, positively charged matrix

**salient property:** much heavier than electrons (about 8000 times heavier)

<sup>19</sup> For a nice picture of Rutherford’s colliding alpha particles, see: <http://www.chemsoc.org/timeline/pages/1911.html>. For his 1911 paper, see <http://www.chemteam.info/Chem-History/Rutherford-1911/Rutherford-1911.html>.



Gold atom**type:** object**structure-of:** *electrons* embedded in large *positively charged matrix***salient property:** cannot strongly deflect alpha particles

Step 2—*Focusing on the alpha particles; first construing.* Zeroing in on the alpha particles and the individual gold atoms, Rutherford cogitates on the experimental results. This gives rise, let us suppose, to this schema, which encapsulates his focus on the odd behavior of the alpha particles:

Alpha Particle Path**type:** relation**relata:** alpha particle, gold atoms**definition:** path followed by alpha particle in relation to gold atoms in gold foil**path description:** the deviation from a straight-line or mildly curved path is very large**salient property:** path contradicts predictions.

This schema is the result of a construing (one occurring, say, in roughly the same time frame as step 1), hence, it is *fully semantical*. Furthermore, it is a distillation of Rutherford's knowledge of the behavior of the alpha particles; it contains no information about *why* the alpha particles behave as they do; indeed at this point, Rutherford is not re-thinking the structure and form of individual gold atoms at all. This is so because (1) the construal Alpha Particle Path occurs early in Rutherford's initial cogitating about the results of his experiment, and hence, (2) Rutherford still conceives of gold atoms in the plum-pudding way.

Meanwhile. Suppose that in long-term memory, Rutherford has this (conceptual) representation of comets and their paths:

Comet Path**type:** relation**relata:** comet, sun**definition:** path followed by comet around sun**path description:** highly eccentric**path types:** elliptic (for orbiting comets); hyperbolic (for one-time comets)**relative sizes:** comet is far smaller than sun.**caused-by:** gravitational pull of sun and speed of comet

Step 3—*Rapid Abstraction and analogical reminding—comets.* This is the Liberman and Trope/rapid abstraction/reminding step. Claim: Alpha Particle Path is instrumental in retrieving Comet-Path. When this occurs, Comet-Path is *rapidly abstracted* via the Liberman and

Trope idea; the abstraction is a function of the semantic distance between Alpha Particle Path (from step 2) and Comet-Path. (This abstracted form of Comet-Path exists in short-term memory;) The relational type remains the same, but the relata—*sun*, *comet*—are abstracted to just two things—*X*, *Y*—the former of which *moves around* the latter. The slots “path types,” “relative sizes,” and “caused-by” are all packed away (packing was defined in “[Rapid abstraction in STRANG](#)”). This leaves the slot “path description.” Crucially, in Comet-Path, the “highly eccentric” filler of the slot “path description” is a quasi-technical term meaning that, for orbiting comets, their ellipse is highly non-circular and flattened, and, for one-time comets, their relevant hyperbolas curve sharply. “Highly eccentric” gets abstracted: it loses its technical status associating it with comets and their orbits. Here is the schema:

Comet Path-Retrieved**type:** relation**relata:** *X*, *Y***definition:** path followed by *X* around *Y***path description:** *X* is highly deflected around *Y*

Step 4—*Rutherford's analogical insight—the second construing.* This is the final analogy building step; it involves a second construing. The abstract Comet-Path-Retrieved is now involved in a second construal process with Alpha Particle Path. Here, too, in accordance with the construing model, some representational change can occur. We thus get our sought-after analogical insight: Alpha Particle Comet-like Path—the construed combination of Comet-Path-Retrieved (from step 3) and Alpha Particle Path (from step 2).

Alpha Particle Comet-like Path**type:** relation**relata:** alpha particle, gold atom**definition:** path followed by an alpha particle in relation to a gold atom**path description:** alpha particle is highly deflected around gold atom**salient property:** path contradicts predictions

Several comments are in order. (1) Alpha Particle Comet-like Path constitutes the *analogical insight* of seeing alpha particles colliding with gold atoms as like comets moving around the sun. (2) Alpha Particle Comet-like Path is the schema derived, via construing, from both Alpha Particle Path and the retrieved and rapidly abstracted version of Comet-Path, namely Comet-Path-Retrieved. It is only approximately the shared structure between these two schemas. More importantly, it did not exist ahead of time, and it was not discovered—it was constructed rapidly in real time using both rapid abstraction and construing.

This is crucial. If my hypothesis is correct, analogy's ability to quickly produce insight and new knowledge is revealed as a dynamic process of abstraction and categorization (construal) put together. (3) Note that a new part of Alpha Particle Comet-like Path is the filler for the slot "path description." The underlined relation in that filler, "highly deflected around," is derived from Comet-Path-Retrieved; the relata are derived from Alpha Particle Path. (4) Alpha Particle Comet-like Path is only the initial analogy. For one thing, it contains *no* slot "caused-by" as Comet-Path did (in long-term memory). This is because at the time Rutherford had this analogical insight, he did not know the cause of the alpha-particles unpredicted and unusual behavior. Most the hard conceptual and technical work initiated by this insight remains to be done. This remaining work also involves analogy and is, I think, properly called *analogical reasoning*. Hence, the four steps above are *not* intended to be an explanation of how Rutherford eventually derived his "solar system" model of the atom. In particular, Rutherford's conclusion that the entire positive charge of an atom was located in its very small center (what we now call the nucleus) is not represented above at all. The history of the development of his model is complex: it took quite a while and involved a lot of cogitation. The analogical insight discussed here occurred at the beginning of all this cogitation. This insight was no doubt central to abandoning the Thomson model of the atom in favor of the more accurate Bohr model.

## Conclusion

My goal here has been to describe analogical insight using several key ingredients. One is rapid abstraction during retrieval via the Liberman–Trope mechanism. Another is construing. The rapid abstraction-construing theory of analogy unites categorization and analogy-making by rendering the latter a species of the former. The conclusions to draw from this view of analogy are that spontaneous analogical reminding produces insight and new knowledge, that it does this in part because of a special kind of representational change, namely, rapid abstraction, that this kind of representational change occurs quickly, at the time the analogy is forged, and that construing is what does the final unification.

Rutherford once said: "In science there is only physics; all the rest is stamp collecting." It is ironic, therefore, that we are using Rutherford's analogical insight, his thoughts, and his cognitive representations—his *mind*—to construct a *scientific* theory about rapid human analogy making, a part of human psychology, neither of which are physics, nor stamp collecting. We can construe Rutherford, having

his analogical insight, as a wayward alpha particle, revealing the nature of the mind. Hopefully, he's not too upset.<sup>20</sup>

## References

- Atkins L (2004) Analogies as categorization phenomena: studies from scientific discourse. Ph.D. dissertation, University of Maryland, College Park
- Barsalou L (2003) Abstraction in perceptual symbol systems. *Phil Trans R Soc Lond B* 358:1177–1187
- Biederman I (1987) Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94:115–147
- Bruner JS (1957) Going beyond the information given. In: Bruner JS, Brunswik E, Festinger L, Heider F, Muenzinger KF, Osgood CE, Rapaport D (eds) *Contemporary approaches to cognition*. Harvard University Press, Cambridge, pp 41–69
- Day SB, Goldstone RL (2009) Analogical transfer from interaction with a simulated physical system. In: *Proceedings of the thirty-first annual conference of the cognitive science society*
- Dietrich E (2000) Analogy and conceptual change, or you can't step into the same mind twice. In: Dietrich E, Markman A (eds) *Cognitive dynamics: conceptual change in humans and machines*. Lawrence Erlbaum, Mahwah, pp 265–294
- Dietrich E (2006) Representation. In: Gabbay DM, Thagard P, Woods J (eds) *Handbook of philosophy of psychology and cognitive science. Handbook of the philosophy of science*, vol 12. Elsevier Science, Amsterdam
- Dietrich E (2008) Toward extending the relational priming model: six questions. *Behav Brain Sci*
- Dietrich E, Markman A, Winkley M (2003) The prepared mind: the role of representational change in chance discovery. In: Ohsawa Y, McBurney P (eds) *Chance discovery by machines*. Springer, Berlin, pp 208–230
- Dixon JA, Dohn MC (2003) Redescription disembeds relations: evidence from relational transfer and use in problem solving. *Memory Cogn* 31:1082–1093
- Doumas L, Hummel J, Sandhofer C (2008) A theory of the discovery and prediction of relational concepts. *Psychol Rev* 115(1):1–43
- Dretske F (1981) *Knowledge and flow of information*. MIT, Cambridge
- Dunbar K (2001) The analogical paradox: why analogy is so easy in naturalistic settings, yet so difficult in the psychological laboratory. In: Gentner D, Holyoak K, Kokinov B (eds) *The analogical mind*. MIT, Cambridge
- Falkenhainer B (1990a) A unified approach to explanation and theory formation. In: Shrager Langley (ed) *Computational models of scientific discovery and theory formation*. Morgan Kaufmann, San Mateo

<sup>20</sup> The research for and work on this paper has benefitted from good conversations with Rick Dale and Clay Morrison. This version also benefitted from good comments made by two careful referees. Previous versions dating back to a distant time have benefitted from conversations with Dedre Gentner, Ken Forbus, Rob Goldstone, Michiharu Oshima, and Clay Morrison. I also thank Art Markman for many discussions on analogy, representation, cognitive science, and a host of other topics. And finally, as always, I thank Chris Fields for decades of exciting, learned, insightful, conversation about everything from analogy to zombies.

- Falkenhainer B (1990b) Analogical interpretation in context. In: Proceedings of the twelfth annual conference of the cognitive science society. Lawrence Erlbaum, Cambridge
- Falkenhainer B, Forbus K, Gentner D (1989) The structure-mapping engine: algorithm and examples. *Artif Intell* 41(1):1–63
- Forbus K, Gentner D, Law K (1995) MAC/FAC: a model of similarity-based retrieval. *Cogn Sci* 19:141–205
- Forbus K, Gentner D, Markman A, Ferguson R (1998) Analogy just looks like high-level perception. Why a domain-general approach to analogical mapping is right. *J Exp Theor Artif Intell* 10:231–257
- French RM (1995) The subtlety of sameness: a theory and computer model of analogy-making. MIT, Cambridge
- Gentner D (1983) Structure-mapping: a theoretical framework for analogy. *Cogn Sci* 7:155–170
- Gentner D (1989) The mechanisms of analogical learning. In: Vosniadou S, Ortony A (eds) *Similarity and analogical reasoning*. Cambridge University Press, Cambridge, pp 199–241
- Gentner D, Colhoun J (2010) Analogical processes in human thinking and learning. In: von Müller A, Pöppel E (Series eds), Glatzeder B, Goel V, von Müller A (Vol. eds) *On thinking: vol. 2. Towards a theory of thinking*. Springer, Berlin
- Gentner D, Wolff P (2000) Metaphor and knowledge change. In: Dietrich E, Markman A (eds) *Cognitive dynamics: conceptual and representational change in humans and machines*. pp 295–342
- Gick ML, Holyoak KJ (1980) Analogical problem solving. *Cogn Psychol* 12:306–355
- Holyoak K, Thagard P (1989) Analogical mapping by constraint satisfaction. *Cogn Sci* 13:295–355
- Holyoak KJ, Thagard P (1995) *Mental leaps: analogy in creative thought*. MIT, Cambridge
- Hummel J, Choplin J (2000) Toward an integrated account of reflexive and reflective reasoning. In: Gleitman LR, Joshi AK (eds) *Proceedings of the twenty-second annual conference of the cognitive science society*. Erlbaum, Mahwah, pp 232–237
- Hummel JE, Holyoak KJ (1997) Distributed representations of structure: a theory of analogical access and mapping. *Psychol Rev* 104(3):427–466
- Hummel JE, Holyoak KJ (2003) A symbolic-connectionist theory of relational inference and generalization. *Psychol Rev* 110(2):220–264
- Indurkha B (1992) *Metaphor and cognition*. Kluwer, Dordrecht
- Indurkha B (1997) Metaphor as change of representation. *J Exp Theor AI* 9(1):1–36
- Indurkha B (1998) On creation of features and change of representation. *J Jpn Cogn Sci Soc* 5(2):43–56
- Keane M (1996) On adaptation in analogy: tests of pragmatic-importance and adaptability in analogical problem solving. *Q J Exp Psychol* 49A:1062–1085
- Keane M (1997) What makes an analogy difficult? The effects of order and structure on analogical mapping. *J Exp Psychol Learn Mem Cogn* 23:946–967
- Keane M, Brayshaw M (1988) The incremental analogy machine: a computational model of analogy. In: Sleeman D (ed) *Third European working session on machine learning*. Pitman, London
- Keane M, Ledgeway T, Duff S (1994) Constraints on analogical mapping: a comparison of three models. *Cogn Sci* 18:387–438
- Kokinov B, Petrov A (2001) Integrating memory and reasoning in analogy-making: the AMBR model. In: Gentner D, Holyoak K, Kokinov B (eds) *The analogical mind*. MIT, Cambridge
- Kuehne SE, Forbus KD, Gentner D, Quinn B. (2000) Category learning as progressive abstraction using structure mapping. In: Proceedings of the twenty-second annual conference of the cognitive science society, pp 770–775
- Kurtz K (2005) Re-representation in comparison: building an empirical case. *J Exp Theor Artif Intell* 17:447–459
- Kurtz K (2007) The divergent autoencoder (DIVA) model of category learning. *Psychon Bull Rev* 14(4):560–576
- Kurtz K, Dietrich E (2006) *Construing categories*. Unpublished manuscript
- Leech R, Mareschal D, Cooper RP (2008) Analogy as relational priming: a developmental and computational perspective on the origins of a complex cognitive skill. *Behav Brain Sci*
- Lieberman N, Förster J (2009) The effect of psychological distance on perceptual level of construal. *Cogn Sci* 33(7):1330–1341
- Lieberman N, Trope Y (2008) The psychology of transcending the here and now. *Science* 322(21):1201–1205
- Lipson J (1981) *Elements of algebra and algebraic computing*. Benjamin/Cummings, Menlo Park
- Markman A, Dietrich E (1998) In defense of representation as mediation (plus several replies to commentaries). *Psychology*. 98.9.48.representation-mediation. <http://www.cogsci.ecs.soton.ac.uk/cgi/psyc/newpsy?9.48>
- Markman A, Dietrich E (2000) In defense of representations. *Cogn Psychol* 40:138–171
- Mitchell M (1993) *Analogy-making as perception: a computational model*. MIT, Cambridge
- Ramscar M, Yarlett D (2003) Semantic grounding in models of analogy: an environmental approach. *Cogn Sci* 27:41–71
- Rapaport WJ (1988) Syntactic semantics: foundations of computational natural-language understanding. In: Fetzer JH (ed) *Aspects of artificial intelligence*. Kluwer, Dordrecht, pp 81–131
- Rapaport WJ (1995) Understanding understanding: syntactic semantics and computational cognition. In: Tomberlin JE (ed) *Philosophical perspectives, vol 9: AI, connectionism, and philosophical psychology*. Ridgeview, Atascadero, pp 49–88
- Rapaport WJ (1999) Implementation is semantic interpretation. *Monist* 82:109–130
- Skorstad J, Gentner D, Medin D (1988) Abstraction processes during concept learning: a structural view. In: Proceedings of the 10th annual conference of the cognitive science society, Erlbaum
- Thagard P, Holyoak K, Nelson G, Gochfeld D (1990) Analog retrieval by constraint satisfaction. *Artif Intell* 46:259–310
- Tsotsos JK (1990) Analysing vision at the complexity level. *Behav Brain Sci* 13–3:423–445
- Tsotsos JK (1995) Behaviorist intelligence and the scaling problem. *Artif Intell* 75:135–160
- Turney P (2008) The latent relation mapping engine: algorithm and experiments. *J Artif Intell Res* 33:615–655
- Usborne M, Lee B (1997) Going beyond the information given: seeing versus seeing as. *Perception* 26
- Wilson D (1983) *Rutherford: simple genius*. MIT, Cambridge
- Winkley M (2006) *Strang: a model of analogical retrieval-based structural change*. Ph. D. dissertation, Philosophy and Cognitive Science Program, Binghamton University, Binghamton, New York
- Wisniewski E (1997) When concepts combine. *Psychon Bull Rev* 4(2):167–183