Overall Similarity Overrides Element Similarity when Evaluating the Quality of Analogies*

Ricardo A. Minervino,[1] Nicolás Oberholzer,[2] and Máximo Trench[3]

[1]*Universidad Nacional del Comahue, Argentina*
minervinor@jetband.com.ar
[2]*Universidad Nacional de Buenos Aires, Argentina*
nicolasoberholzer@hotmail.com
[3]*Universidad Nacional del Comahue, Argentina*
maxtrench@gmail.com

Dominant computational models of analogical reasoning (e.g., SME and LISA) consider that two facts or situations are more analogous as the similarity between corresponding propositional elements increases. We report the results of two experiments demonstrating that when people judge the quality of an analogy, the similarity between matched elements is overridden by another type of similarity that implies comparing the meaning of whole propositions. In Experiment 1, participants received a base fact followed by two structurally identical target facts. Whereas in one of them propositional elements resembled their counterparts in the base, in the other they did not, but the meaning of the whole proposition resembled that of the base. Participants chose as more analogous the targets maintaining this second type of similarity. In Experiment 2, participants received a base cause followed by an effect, and were told that such effect reoccurred later as a consequence of an analogous cause. Participants had to decide which of two structurally identical facts was the

---

cause of the target effect. Again, participants based their choices on overall similarities, passing over similarities between propositional elements, but in a more ecologically valid task that involves comparing systems of relations. We conclude with some intuitions about the mechanisms underlying how people assess the quality of an analogy, and discuss their implications for future theories of analogical thinking.

Key words: *analogy, similarity, relational categories*

Drawing an analogy consists in recognizing that the entities of two situations are organized by similar systems of relations (Gentner, 1983; Holyoak, 1984). Across activities as diverse as persuasion, problem solving, decision making or explanation, people often have to assess to what extent two situations are analogous, frequently with the purpose of deciding if knowledge about one of such situations (typically the *base analog*: BA) can be sensibly projected onto the other (the *target analog*: TA) (Gentner, Holyoak, & Kokinov, 2001; Holyoak & Thagard, 1995).

The present study set forth to determine if the semantic mechanisms incorporated by dominant computational models of analogical mapping and inference (i.e., SME, Falkenhainer, Forbus, & Gentner, 1989; Forbus, Ferguson, & Gentner, 1994; and LISA, Hummel & Holyoak, 1997, 2003) are adequate to account for the judgments people make about the degree to which two situations can be considered analogous. One way of examining how the semantic constraints incorporated in computer models of mapping and inference generation affect how they judge the quality of analogies would be to present them with a BA followed by two TAs varying along a relevant semantic dimension, and observe which of the competing TAs the system considers to be a better match to the BA (e.g., Gentner & Kurtz, 2006; Gentner, Rattermann, & Forbus, 1993; Hummel & Holyoak, 1997). For example, to assess the effect of object similarity on such judgments, one could present the above systems with a BA (e.g., *Peter mows the grass*) followed by two alternative TAs: one of them maintaining object and relational similarity with the BA (TA1: *Marcus cuts the weeds*), and the other maintaining only relational similarity with the BA (TA2: *Samuel cuts the meat*). If a given program takes into account object similarities in

its judgments of the quality of an analogy, this would become apparent in a preference for TA1 over TA2 as a match to the BA. Since this task also turns out to be rather meaningful for human reasoners, it allows a straightforward comparison between the way humans and computational models judge the quality of analogies, thus serving to assess the psychological adequacy of the semantic assumptions embraced by extant computational models.

In the next section we employ the above criterion to analyze how the semantic constraints incorporated by dominant computational models determine their judgments of analogical quality. Then we flesh out an analysis of the limits that inhere current model's treatment of semantics as similarities between paired elements, and present results from two experiments demonstrating that human reasoners sometimes pass over these element-to-element similarities in favor of more holistic criteria. Finally, we conclude with some intuitions about the mechanisms underlying these criteria. To preview, we will suggest that the quality of an analogy is sometimes decided in terms of whether or not two situations can be assigned to a shared relational category.

1. The standard approach to the role of semantics when assessing the quality of analogies

The structure-mapping theory (Gentner, 1983, 1989; Gentner & Markman, 1997, 2005) and the multiconstraint theory (Holyoak & Thagard, 1989; Hummel & Holyoak, 1997, 2003) have dominated the discussion and computational modeling of analogical reasoning. As their computational accounts of the mapping and evaluation stages of analogical reasoning are representative of the main proposals that have appeared since the 80s, we will take them as the main exemplars of what we will henceforth refer to as the *standard approach* to these subprocesses (see Gentner & Forbus, 2011, and French, 2002, for reviews of computational models of analogical reasoning).

The structure-mapping theory postulates that knowledge is represented in propositional form, and distinguishes between: a) *entities*: single elements that stand for objects and individuals; b) *attributes*: unary predicates representing properties of entities; c) *first-order relations*: multiplace

predicates that link two or more entities; and d) *higher order relations*: predicates that link relations themselves. According to this theory, two situations are judged to be analogous when their entities are organized by semantically similar relational structures—systems of relations governed by higher order relations—and satisfy a number of syntactic constraints (Gentner, 1989; Gentner & Markman, 1997, 2005). SME (Falkenhainer, Forbus, & Gentner, 1989), the core computational implementation of the theory, is a symbolic system that takes as inputs propositional descriptions of the BA and the TA, and finds the maximal (i.e., largest and deepest) coherent relation and entity match between the two, leaving aside isolated relations. Within the SME architecture two elements are allowed to be mapped only if they satisfy the following initial conditions: a) *formal identity*: elements must be of the same formal type (relations, entities, *n*-place relations, etc.); and b) *semantic similarity for relations and entity properties*: relations and properties can be mapped only if they are similar in meaning. Once all local matches are generated, the program incrementally merges them into a few global mappings. Such mappings are *structurally consistent*, that is, they satisfy the following constraints: a) *parallel connectivity*: if two predicates are put in correspondence, their arguments must also be mapped; and b) *one-to-one mapping*: each element in the BA must map to at most one element in the TA and vice-versa. SME uses the established mappings to suggest hypotheses about the TA. Finally, each global mapping is given an evaluation based on the number of local matches, the depth of the system of matches (the *systematicity principle*; Gentner, 1983), and the degree of similarity of the matched relations.

Suppose that SME receives a BA stating that *John loves Mary* together with two alternative TAs, one stating that *Peter likes Susan* (TA1) and another stating that *Richard fears Beth* (TA2). Given that both alternatives could equally satisfy syntactic principles like one-to-one mapping and parallel connectivity, the program would decide between them based solely on semantic grounds. The rerepresentational mechanisms incorporated to SME in successive extensions of the program were engineered to discover underlying identities between similar but initially non-identical relational predicates. Through a mechanism called *minimal ascension* (Falkenhainer, 1990), the system would search for the least abstract common superordinate

of two relations along an IS-A hierarchy of relations. Whenever two TAs compete at matching a BA, the TA requiring fewer steps to find a shared superordinate will be considered to be more analogous. Applying minimal ascension to the above example, SME would prefer to match *loves* to *likes* rather than to *fears*, on the grounds that the superordinate concept of *love* and *like*, say, "feel affection", can be found without escalating to rather abstract concepts along the conceptual hierarchy (say, "feel something"). A more recent strategy incorporated by SME to discover identities between initially non-identical relations is *decomposition* (Yan, Forbus, & Gentner, 2003), a mechanism that involves breaking down relational predicates into the subcomponents that encode their meaning, in order to find identity matches between these subcomponents. Applying decomposition to the BA and the TA1 presented above could result in:

BA: LOVE (John, Mary) → IMPLY [LOVE (John, Mary), FEEL AFFECTION (John, Mary)]
TA1: LIKE (Peter, Susan) → IMPLY [LIKE (Peter, Susan), FEEL AFFECTION (Peter, Susan)].

If the system were confronted with two competing TAs for a given BA, it would prefer the TA for which the emerging identity arises at a lower level of abstraction. Returning to the above example, given that decomposition would not reveal identities between *loves* and *fears* (or it would find a rather abstract one), SME enhanced with this rerepresentation mechanism would also prefer matching *John loves Mary* to *Peter likes Susan* rather than to *Richard fears Beth*.

The multiconstraint theory of Holyoak and colleagues (Holyoak & Thagard, 1989; Hummel & Holyoak, 1997) conceives mapping as determined by a conjunction of syntactic, semantic and pragmatic constraints (the structure-mapping theory has contended that pragmatic factors operate either before or after—but nor during—the mapping and inference generation stages; see Gentner, 1989; Gentner & Markman, 1997). LISA (Hummel & Holyoak, 1997, 2003), the current computational implementation of the multiconstraint theory, is a hybrid system that combines the semantic flexibility of connectionist architectures with the

sensitivity to structure provided by symbolic models. The core of LISA's architecture is a system for representing dynamic role-filler bindings in working memory (WM) and encoding these bindings in long-term memory (LTM). When a proposition unit (P) gets activated, it propagates top-down activation to subproposition units (SPs) representing bindings between each of the case roles of the proposition and its corresponding filler. During the lapse while each SP unit remains active, it transfers top-down activation to two independent structure units representing a case role and its filler. These two types of structures—which represent the lower level in the structural hierarchy—in turn, activate a collection of semantic units representing their meaning. As different SP units within a proposition inhibit each other and inhibit themselves to inactivity once activated, they become active in an alternating fashion. Therefore, when a proposition such as *John loves Mary* is selected, the semantic primitives of *lover* (e.g., *emotion1*, *positive1*, and *strong1*) fire in synchrony with the semantic primitives of *John* (e.g., *human*, *male* and *adult*), while units representing the *beloved* role (e.g., *emotion2, positive2* and *strong2*) fire in synchrony with units representing *Mary* (e.g., *human*, *female* and *adult*). When the semantic primitives of a given role-filler binding in the target are selected to fire in WM, predicate, object and SP units from one or various sources compete in responding to this array as a function of the extent to which their semantic units overlap. Since the mechanism through which predicate and object units get activated is essentially the same (i.e., via shared semantic primitives) the semantic similarity constraint affects objects and relations evenly.

In LISA, syntactic constraints are enforced by sets of excitatory and inhibitory links. Within an analog, units of different hierarchy are linked by symmetric excitatory connections, whereas units of the same level share symmetric inhibitory links. In this way, when a predicate and an object unit in the source respond to patterns of activation in WM, they activate SP and P units above them, all of which tend to inhibit other units of the same type, thus enforcing a one-to-one mapping constraint. Once a P unit in the target has activated a corresponding P unit in the source, parallel connectivity is enforced by top-down activation of the structure units below them. Mapping hypotheses in LISA are connection weights created to encode associations in LTM between structure units of the same type

across different analogs. The system increments the positive weights of these connections as a result of their temporal co-activation. The one-to-one constraint is further enforced in LISA through an algorithm that responds to an increment in the weight of a mapping connection between two units by decreasing the weight of all other connections leading to such units. LISA can put in correspondence two similar but non-identical relations and objects. Consider once again matching *John loves Mary* to *Peter likes Susan* versus *Richard fears Beth*. The SP for John-as-lover will activate the semantic units of *John* (e.g., *human*, *male* and *adult*) and *love1* (*emotion1*, *positive1* and *strong1*). This pattern will excite object and predicate units in the sources, which will compete to become active. *Human, male* and *adult* will excite *Peter* and *Richard*, whereas *human* and *adult* will excite *Susan* and *Beth*. In this competition, *Peter* and *Richard* will become equally active, inhibiting *Susan* and *Beth*. Based on their semantic overlap alone, LISA begins to act as if *John* corresponded to either *Peter* or *Richard*. At the same time, *emotion1*, *positive1* and *strong1* will excite the predicate unit *likes*, but only *emotion1 and strong1* will excite *fears1*. As *likes1* will inhibit *fears1*, LISA begins to act as if *loves1* corresponded to *likes1*. Because *likes1* is more active than *fears1*, the SP *Peter+likes1* will receive more bottom-up input—and therefore become more active—than the SP *Richard+fears1*. As SPs excite the P units to which they belong, the unit for *Peter likes Susan* will become more active than the unit for *Richard fears Beth*. Hence, LISA concludes that *John loves Mary* is more analogous to *Peter likes Susan* than to *Richard fears Beth*. The SP mappings allow LISA to resolve the semantically ambiguous John-to-Peter versus John-to-Richard mappings. SPs feed activation back to their predicate and object units, giving *Peter* an edge over *Richard*. Now, LISA concludes that *John* corresponds to *Peter* rather than to *Richard*. Analogous operations will cause LISA to conclude that *Mary* corresponds to *Susan* rather than to *Beth*, and that *loves2* corresponds to *likes2*.

In sum, whereas the structure-mapping theory posits that only relations count when evaluating the quality of an analogy, the multiconstraint theory contends that both objects and relations contribute to quality evaluations—a position also shared by authors such as Keane, Hackett and Davenport (2001), and Larkey and Love (2003).

A few studies sought to investigate whether or not human reasoners take object matches into account during the evaluation of analogical relatedness. The most thorough-going investigation of how people judge the quality of analogies across different kinds of matches is the "Karla the Hawk" series conducted by Gentner et al. (1993). Upon receiving pairs of stories that maintained different kinds of commonalities (including analogies and literal similarities), participants were asked to rate each pair for similarity and for inferential soundness—an operationalization of the quality of an analogy that consists of assessing to what extent one story could be used to draw inferences about the other. Regarding similarity judgments, literal similarities were judged as slightly more similar than analogies, thus showing a small but significant contribution of object similarities to overall similarity. However, in contrast to similarity ratings, soundness ratings were not affected by object similarities.

A limitation of the above study concerns the extent to which judgments of analogical soundness really inform how people perceive the quality of an analogy. From our point of view, Gentner et al.'s (1993) instructions were ambiguous as to whether they asked about the straightforwardness with which inferences can be derived from the analogy (i.e., soundness), or the plausibility of these inferences within the target domain. We tend to agree with Gentner and Kurtz (2006) on the convenience of probing judgments of analogical quality with more direct questions such as "To what extent do you consider these situations analogous?" Gentner and Kurtz (2006, Experiments 1 and 2) asked participants to give timed answers to whether a BA (e.g., *John bought the candy*) was analogous to TAs in which the base relation (represented by a verb) or the base object (represented by a noun) was replaced by verbs or nouns of varying degrees of semantic similarity. For the above BA, examples of verb substitutions were *John purchased the candy* (synonymous), *John took the candy* (near) and *John stepped on the candy* (far). Likewise, examples of noun substitutions were *John bought the sweets* (synonymous), *John bought the sandwich* (near) and *John bought the bookshelf* (far). Judgments of the quality of the analogies were highly sensitive to the degree of relational match and much less sensitive to the degree of object match. As relational similarity diminished, analogy judgments declined from nearly universal acceptability for synonymous

verbs to nearly universal rejection for far verbs. The pattern for nouns was quite different. There was no drop-off in analogical acceptability from synonymous nouns to near nouns; not until the lowest level of object similarity did analogical acceptability show a significant (but small) drop. Even at this lowest level, analogical acceptance remained the dominant response, at roughly 80% across Experiments 1 and 2. In sum, studies addressing the relative weight of relations and objects on the evaluation of analogies show that relations count much more than objects.

2. Shortcomings of the Dominant Theories in their Treatment of Semantics

Despite their differences in terms of architectures, mechanisms and postulated representations, the revised computational models of analogical evaluation will consider two situations to be more analogous as the similarity between corresponding elements (relations and eventually objects, depending on the model) increases. However, this treatment of semantics seems inappropriate in those particular cases where the interaction between the elements composing the analogs invites descriptions that severely depart from the meaning of any of these elements considered in isolation. Sticking to Gentner and Kurtz's (2006) example of near verb substitutions, the "standard" meaning of the verbs *buy* and *take* (i.e., when they are considered in isolation) is preserved when they are bound to the argument *candy*, yielding the propositions *John bought the candy* vs. *John took the candy*. As their meanings are preserved within both propositions, their "standard" similarity is also maintained. However, people will probably refuse to consider *John bought a purse from the old lady* as analogous to *Phil took a purse from the old lady*, on the grounds that the second action—but not the first—represents a case of robbery.[1]

---

[1] Gentner and France (1988) carried a study in which participants paraphrased sentences that combined verbs and nouns with varying degrees of semantic strain (e.g., *The lizard worshipped*). An analysis of such paraphrases revealed that verbs alter their meaning more than nouns do (see also Kesten & Earles, 2004). In the case we are considering (and in the materials of our experiments) the noun does not activate a particular meaning of the verb that applies to it—nor does it promote

The tasks we have been employing to show the role of semantics on analogical thinking (i.e., one BA and two competing TAs) could also serve to illustrate how the computation of element similarities can lead current models to misjudge the quality of an analogy. Consider, for example, a BA stating that *John bought a perfume for Mary*, followed by *Peter lent his deodorant to Susan* (TA1), and *Richard wrote a poem to Beth* (TA2). Despite the existing similarities between corresponding elements of the BA and the TA1 (*buy* is similar to *lend* but not to *write*; *perfume* is similar to *deodorant* but not to *poem*), TA2 could be considered more analogous to the BA, on the grounds that a man giving a perfume to a girl and a man writing a poem to a girl are both cases of, say, *seduction* or *expressing love*, whereas lending a deodorant to a girl should not be regarded as a case of *seduction* or *expressing love*—in fact, in can even be considered rude.

Confronted with the above task, LISA would prefer the mapping between the BA and the TA1, in response to the higher degree of similarity between their corresponding relations (*buy* and *lend*) and objects (*perfume* and *deodorant*). The alternative mapping implies matching less similar relations (*buy* and *write*) and objects (*perfume* and *poem*). When the SP for *John+buyer* fires in LISA's WM, it will activate units for *John* and *buy1*, transferring top-down activation to their semantic units (e.g., *human*, *male*, *adult* and *action1*, *pay1*, *transfer1*, respectively). This pattern will excite units in the TAs, which will compete for activation. *Human*, *male* and *adult* will transfer more activation to the units corresponding to *Peter* and *Richard* in the targets, than to the units corresponding to *Susan* or *Beth*. At the same time, *transfer1* and *action1* will excite the predicate unit *lend1*, but only *action1* will excite *write1*. As *lend1* will inhibit *write1*, LISA begins to act as if *buy1* corresponded to *lend1*. Analogous operations will cause LISA to act as if *buy2* corresponded to *lend2*. Accordingly, triggered by the SP *perfum+object* that was built for the BA, LISA will map *perfume* to *deodorant* and *object* to *object*. Given that *lend1* is more active

----

an extended use of it (e.g., a metaphorical one)—but it activates a concept whose meaning is entirely different from the verb on which it originates. For instance, *John put poison in Mary's soup* prompts the concept "murder attempt", whose meaning does not include the concept of *putting* as an informative component.

than *write1*, the SP *Peter+lender* will receive more bottom-up input than the SP *Richard+writer*. Analogous support will be obtained by the target SP units *Susan+lendee* and *deodorant+object*. Finally, as SPs excite the P units to which they belong, the unit for LEND (*Peter, Susan, deodorant*) will become more active than the unit for WRITE *(Richard, Beth, poem)*. Given BUY (*John, Mary, perfume*) as a BA, LISA will thus consider LEND (Peter, Susan, deodorant) as more analogous than WRITE (Richard, Beth, poem).

Even though SME's evaluation of analogical quality would only take into account semantic similarities between relations, it would still choose the TA1 as more analogous. In so doing, the program would need to apply, for example, the rerepresentation mechanism of semantic decomposition proposed by Yan, Forbus and Gentner (2003) described above. Applying decomposition to BA and TA1 could result in:

> BA: BUY [(John, Mary, perfume) → IMPLY (BUY (John, Mary, perfume), TRANSFER (John, Mary, perfume)].
> TA1: LEND [(Peter, Susan, deodorant) → IMPLY (LEND (Peter, Susan, deodorant), TRANSFER (Peter, Susan, deodorant)].

In this case an identical semantic element (*transfer*) is found within the meanings of both predicates. In contrast, applying decomposition to BA-TA2 would fail to find a meaningful identity between the relations BUY and WRITE—or it would find a very abstract one—, probably leading the system to reject the analogy between the BA and TA2. Hence, SME augmented with these rerepresentational methods would also map the BA to the TA1.[2]

---

[2] In this example, the particular way of breaking down relations and objects into their semantic primitives could have been resolved in many other ways, favoring perhaps the alternative mapping. For example, had *buy-for* and *write-for* (instead of *buy* and *write*) been coded with semantics such as *generous* or *thoughtful*, then LISA would consider *write-for* as a more viable match for *bought-for*. However, in deciding how to encode the meaning of elements one must try to avoid committing 20/20 hindsight, that is, tailoring the initial coding of the concepts so as to favor any particular interpretation of the analogy. What is clear is that our encoding (be it more or less accurate) should result in evaluations of element similarities that

The programs we have been describing could be extended to generate a common descriptor for the BA and TA1 by means of retrieving and combining the superordinates of the matched propositional elements (e.g., yielding a description such as *a man gives toiletries to a girl*). However, we fail to see how they could give rise to a descriptor that would enable the system to accept the TA2 as analogous to the BA (e.g., *a man is trying to seduce a woman*). As stated, what leads all these programs to choose the TA1 is the fact that they are considering similarities between corresponding base and target propositional elements in isolation, disregarding any similarities that could result from considering the meaning of the propositions taken as wholes, and which originate in complex interactions between the agents of an action, the objects that are involved in such action, and the patients to whom the action is exerted. While in LISA this similarity is reflected through the co-activation of semantic primitives in WM, in SME this similarity can be revealed via rerepresentation mechanisms, such as a minimal ascension or semantic decomposition. We will use the term *element similarity* to refer to similarities between propositional elements that are compared in an isolated form. In contrast, we will use *overall similarity* to name those resemblances maintained by propositions considered as wholes, and that cannot be derived via combining the common superordinates of the mapped elements. As the example above suggests, element similarity is not always the best way to assess the quality of an analogy, and can even be misleading when relatively superficial pairings compete with more adequate descriptions of the analogs. Our first experiment was developed to show that two analogs sharing element similarities could frequently be considered less analogous than two analogs that do not maintain element similarities, but maintain overall similarities instead.

---

reflect people's judgments. When coding this particular example, we had in mind that participants of an independent group (see Experiment 1) had judged that *buy* is more similar to *lend* than to *write*.

3. Experiment 1

We presented participants with a BA followed by two alternative TAs, with the instruction to indicate which of them they considered as more analogous to the BA. The TA1s maintained only element similarities with their respective BAs, thus allowing the extraction of a common descriptor by means of combining the superordinate concepts shared by the matched elements. In contrast, TA2s maintained only overall similarity with the BA, which means that the descriptor subsuming both analogs could not arise from the combination of common superordinates. Within each set, both TAs were formally identical to the BA, that is, a one-to-one mapping satisfying parallel connectivity could be built between each of them and the BA. As was analyzed, in a task with this structure, SME and LISA would choose TA1s as more analogous to the BAs, based on the higher degrees of element similarity maintained by their matched relations and objects. Our goal was to determine whether participants would also prefer to match the BA to TAs maintaining this type of similarity, or if they instead preferred pairing the BAs with TAs maintaining overall similarities.

3.1. Method
3.1.1. Participants
Thirty undergraduate students of Psychology at University of Buenos Aires took part in the experiment in partial fulfillment of a course requirement.

3.1.2. Design and Procedure
The independent variable was the degree of element similarity between base and target (high vs. low), a within-subjects variable. The dependent variable was the chosen TA. Materials were presented in written form and participants were tested individually. For each of the trials, participants had to read a first analog (the base one), consisting of a sentence that described a simple fact. Afterwards, they were presented with two other facts (the target ones), and they were asked "Which of these two facts do you find more analogous to the first?"

Table 1. Materials used in Experiments 1 and 2

	Base and target situations	Common Effect*
Set 1		His wife got really happy
Base:	Lucas decided to *go* to *spinning classes*	
Target 1:	Lucas decided to *attend relaxation lessons*	
Target 2:	Lucas decided to *give up fast food*	
Set 2		The teacher smiled
Base:	Sammy *bought a perfume* for a girl in his class	
Target 1:	Sammy *lent a deodorant* to a girl in his class	
Target 2:	Sammy *wrote* a *poem* to a girl in his class	
Set 3		Her friends were surprised
Base:	Delores *hung garlic* on the *door*	
Target 1:	Delores *planted chives* in the *alley*	
Target 2:	Delores *lit* a *candle* in the *basement*	
Set 4		His wife hired a detective to follow him
Base:	Jimmy was *chatting* with his *ex girlfriend* from his *band*.	
Target 1	Jimmy was *talking* to his *ex partners* in the *gang*	
Target 2	Jimmy was *photographing* his *pupils* from *college*	
Set 5		His son thought it was a good idea
Base:	The dictator had his *servant try* his *meal*	
Target 1:	The dictator had his *cook experiment* with *oysters*	
Target 2:	The dictator had his *double attend* the *parade*	
Set 6		His mom disagreed
Base:	Louis *sent soup cans* to the *school*	
Target 1:	Louis *brought* a *turkey sub* to the *club*	
Target 2:	Louis *bought* a *soccer ball* for the *orphanage*	

Note. *Target effects were only included in Experiment 2

3.1.2. Materials

The TA1s were built replacing one or two verbs and one or two nouns in the BAs by verbs and nouns with very similar meaning. For example, in Set 6 of our materials (see Table 1), the BA was *Louis sent soup cans to the school* and the TA1 *Louis brought a turkey sub to the club*. As can be observed, base elements were replaced by similar elements: *send* by *bring*, *soup cans* by *turkey sub*, and *school* by *club*. In contrast, in the TA2 the

same base verbs and nouns were replaced by verbs and nouns with less or no similar meaning. In the set under consideration, the TA2 was *Louis bought a soccer ball for the orphanage*. As can be seen, *send* was replaced by *buy*, *soup cans* by *soccer ball*, and *school* by *orphanage*. In this way, for the BA-TA1 pair only a rather abstract and vapid descriptor could be constructed from element similarities (in this set, *a boy carried food to an institution*), while in the BA-TA2 pair a more meaningful descriptor could be postulated, based on overall similarities instead of element similarities (in this set, *donation*).

Participants received six critical trials and six fillers. To prevent participants from inducing an association between the presence of element similarities and the absence of overall similarity, the filler sets were built such that the TAs either shared both overall and element similarity with the BA or neither of them. For instance, for a BA like *The boy wiped the bedroom*, while the TA1 would state *The boy mopped the bathroom* (BA and TA1 are instances of *cleaning a room*—a descriptor based on element similarity—as well as of *helping with housekeeping*—a descriptor based on overall similarity), the TA2 would state that *The boy stole money* (a situation that does not maintain any kind of similarity with the BA). The order of presentation for the 12 trials and of TA1 and TA2 within each trial was counterbalanced.

To gather independent measures of the degree of similarity between propositional elements to be mapped, we asked an independent group of 30 students (taken from the same population) to rate the similarity between the BA elements (e.g., *soup cans*) and the corresponding elements in the TA1 (*turkey sub*) and the TA2 (*soccer ball*). This group received a list with 32 pairs of concepts (some of them were pairs of objects and some of them pairs of relations) to be evaluated using a 5-point Likert scale ranging from 1 (no similarity) to 5 (high similarity). The order of presentation of the pairs was counterbalanced, and in all of the cases the BA-TA1 pairs were separated from the BA-TA2 pairs of the same set by at least three interleaving pairs taken from different sets.

In order to compare participants' behavior in the experimental group against SME's assumptions (this program takes into account only the similarity between matched relations at evaluating the quality of an

analogy), the first analysis of the scores provided by the similarity-rating independent group was carried out comparing the semantic similarity between the base relation and the TA1 relation against the similarity between the base relation and the TA2 relation. For each set of materials, Wilcoxon Signed Ranks Tests confirmed that the relation in the BA was rated as more similar to that in the TA1 than to that in the TA2, with median [quartiles] ratings of similarity to base relations being 4 [4 4] vs. 1 [1 2], $Z = -4.872$, $p < .001$ (Set 1); 4 [3 4] vs. 2 [1 2], $Z = -4.849$, p < .001 (Set 2); 3 [2.75 3] vs. 1 [1 2], $Z = -4.738$, $p < .001$ (Set 3); 4 [4 5] vs. 2 [2 3], $Z = -4.875$, $p < .001$ (Set 4); 4 [3.5 4.125] vs. 2 [2 2.5], $Z = -4.864$, $p < .001$ (Set 5), and 4 [3 4] vs. 2 [2 2.25], $Z = -4.681$, $p < .001$ (Set 6). Our second analysis of the similarity scores provided by the independent group was intended to compare participants' behavior in the experimental group against LISA's assumptions[3] (this program takes into account both relation and object similarities at assessing the quality of analogies). It was essentially the same as in the previous case, except for the fact that the statistical analysis corresponding to each of our sets of materials was now fed with each participant's average response to the high element similarity pairs (i.e., an average between the similarity scores given to the relation and object pairs extracted from comparing BA against TA1) vs. each participant's average response to the low element similarity pairs (i.e., an average between the similarity scores given to the relation and object pairs extracted from comparing BA against TA2). In set 6, for example, whereas each participant's relation+objects similarity between the BA and TA1 was obtained averaging the similarity scores given to the pairs *send–bring*, *soup can–turkey sub*, and *school–club,* the relation+objects similarity between the BA and the TA2 was obtained averaging the similarity scores given to the pairs *send–buy*, *soup can–soccer ball,* and *school–orphanage.* For each of our sets of materials, Wilcoxon Signed Ranks Tests confirmed that

---

[3] In spite of the fact that the empirical studies reviewed (Gentner et al., 1993; Gentner & Kurtz, 2006) tended to show that only relations, and not objects, are relevant to the task of evaluating the quality of an analogy, we have remained open to the possibility that the latter have a role as well, as asserted by the multiconstraint theory (Holyoak & Thagard, 1989).

Table 2. Target choices, Experiment 1

Set	TA1	TA2	χ^2 (1, N =30)
1	30%	70%	9.60[b]
2	20%	80%	21.60[b]
3	47%	53%	.27
4	33%	67%	6.67[b]
5	27%	73%	13.07[b]
6	10%	90%	38.40[b]

Note. TA: target analog; [b] Significant at .01 level

relations and objects in the BA were rated as more similar to those in the TA1 than to those of the TA2, with the respective median [quartiles] ratings of similarity being 3.5 [3.375 4] vs. 1.5 [1 1.625], $Z = -4.829$, $p < .001$ (Set 1); 4 [3.5 4.5] vs. 2 [1.5 2], $Z = -4.811$, p < .001 (Set 2); 3.667 [3 3.667] vs. 1.667 [1.333 1.667], $Z = -4.834$, $p < .001$ (Set 3); 3.667 [3.333 4] vs. 2 [1.667 2.083], $Z = -4.825$, $p < .001$ (Set 4); 3.75 [3.5 4] vs. 2 [1.75 2.25], $Z = -4.808$, $p < .001$ (Set 5), and 3.333 [3 3.667] vs. 1.667 [1.667 2], $Z = -4.827$, $p < .001$ (Set 6).

3.2. Results and Discussion

Table 2 shows the percentages of participants that chose, for each of the critical trials, TA1 or TA2. Data show that even though propositional elements of the BA were rated as more similar to their corresponding elements in the TA1 by the independent group, participants of the experimental group chose the TA2 as more analogous to the BA in five of the six critical sets (we found no trend in the remaining set). In sharp contrast with LISA or SME, participants' assessments of the quality of analogies seemed to pass over element similarities, favoring more meaningful descriptors than those that can be generated by means of combining the common superordinates of the matched elements.

As in several works studying the evaluation of the quality of analogies (e.g., Bassok & Medin, 1997; Gentner & Kurtz, 2006), our stimuli consisted in first-order propositions. However, the concept of analogy refers, *stricto sensu,* to a comparison between two systems of relations (Gentner, 1989; Holyoak & Thagard, 1995). In addition, the task of evaluating the

quality of analogies is removed from the type of purposeful activities in which analogical reasoning routinely participates (e.g., problem solving, argumentation, explanation, etc.). In Experiment 2 we sought to extend our findings using more natural tasks, as well as materials that comprise mappings between systems of relations.

4. Experiment 2

A central function of analogical reasoning consists of identifying the causes of target situations. Consider, for example, the case of a reasoner who has to choose among competing explanations for why a European country was led to bankruptcy. If the reasoner is familiar with a similar crisis undergone by another European country (and whose cause is known), he or she can identify a plausible cause for the target situation via deciding which of the target candidate causes is more analogous to the cause that gave rise to the familiar situation. In a similar way, participants of Experiment 2 were presented with a BA consisting of a base cause (BC) and its effect. They were told that the base effect later reoccurred as a consequence of a cause *analogous* to the BC, and were presented with two alternative causes (TC1 and TC2) to choose from. The BC could be evenly matched with either of the two TCs under the formal and pragmatic considerations included by programs like SME and LISA. However, propositional elements of the BC were more similar to their corresponding elements in the TC1 than in the TC2.

In this way, the task used in Experiment 1—judging which of two TAs was more analogous to a BA—shifted to one of deciding which of two facts (both candidate arguments of a second-order causal relation) was more analogous to the BC. For each of our sets of materials, participants in a control group received the TCs but not the BA, and were asked to guess which of the two TCs could have generated the target effect. If target-only participants were less inclined than participants in the analogy condition to choose the TC that maintained overall similarity with the BC, then preferences among experimental participants should be attributed to the influence of the BA, and not to a higher intrinsic plausibility of TC2 as a cause for the target effect. We were interested in determining if, contrary
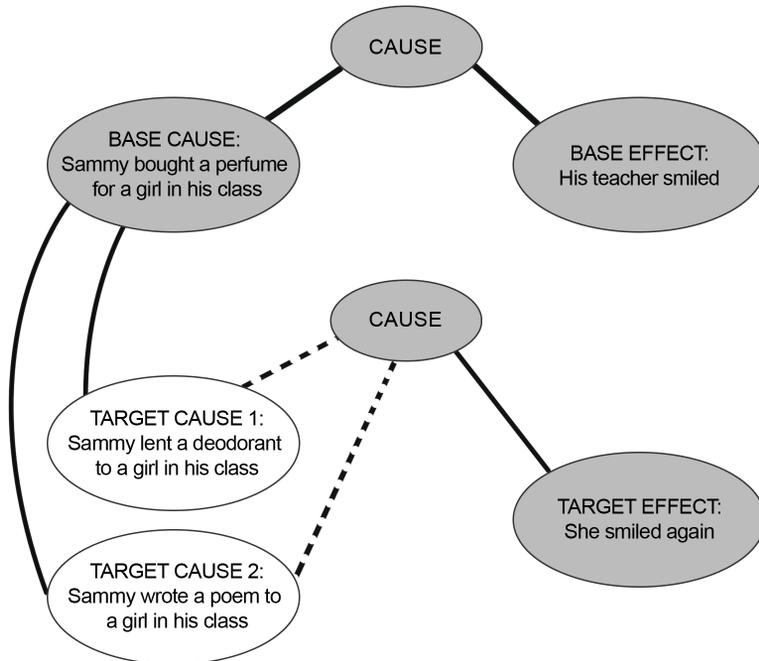
Fig. 1. A representation of the task used in Experiment 2, with an example of materials. The ovals in the upper part represent the system of relations of the BA, which consists of the known cause of certain effect. The grey ovals in the lower part of the figure represent the information given to participants about the TA, which in all cases was the fact that the target effect reoccurred later, for a cause analogous to that of the BA effect (e.g., participants are told that *at a later time the teacher smiled again, for a reason that was analogous to what caused her to smile the previous time*). Finally, the white ovals represent two possible causes for this second effect, from which participants had to choose the one they considered most likely.

to the criteria embedded in programs like SME or LISA, participants of the analogy condition would be more prone than participants in the target-only condition to choose TC2 as the more likely cause of the target event.

4.1. Method

4.1.1. Participants

Sixty students of Psychology at University of Buenos Aires took part in the experiment in partial fulfillment of a course requirement.

4.1.2. Design and Procedure

The independent variables were element similarity between the BC and the TCs (high vs. low), a within-subjects variable, and condition (analogy vs. target-only), a between-groups variable. The dependent variable was the chosen TC. Materials were presented in written form. Participants were tested individually. After reading each BC followed by the action it provoked (i.e., the effect), participants in the analogy group were told that this last action reoccurred some time later as a consequence of an analogous cause. After that, participants in the analogy condition were asked to choose the more likely explanation for such reoccurrence: "Assuming that what caused this second event was *analogous* to what caused the first event, which of these two events would you choose as the possible cause of the second event?" The two candidate TCs were then presented. Participants in the target-only group were presented with the target effect followed by a question to choose which of the alternative facts that were presented afterwards they would consider as the most likely cause for such situation: "Following your subjective criteria, which of these two events would you choose as the possible cause of this event?"

4.1.3. Materials

Participants were presented with six critical trials and six filler trials, which were identical to those of Experiment 1, except for the fact that the BCs—which corresponded to the BAs of Experiment 1—were now followed by an effect (see Table 1). Keeping with the critical set provided as an example in Experiment 1, the BA stated that *Louis sent soup cans to the school* (BC), *causing his mother to disagree* (effect). Right after receiving the BC and its outcome, participants were told that the same outcome (Louis' mother's disagreement with his behavior) reoccurred later as an effect of an analogous cause: "At a later time, his mother disagreed with him as an effect of an analogous fact". The candidate TCs for this reoccurrence

Table 3. Target choices in the analogy group and the target-only group, Experiment 2

Set	Analogy group			Target-only group			Assoc.
	TC1	TC2	$\chi^2(1, N = 30)$	TC1	TC2	$\chi^2(1, N = 30)$	$\chi^2(1, N = 60)$
1	33%	67%	6.67[b]	60%	40%	2.40	4.29[a]
2	27%	73%	13.07[b]	57%	43%	1.07	5.55[a]
3	23%	77%	17.07[b]	50%	50%	.00	4.27[a]
4	43%	57%	1.07	67%	33%	6.67[2]	3.30
5	17%	83%	26.67[b]	63%	37%	4.27[1]	13.61[b]
6	23%	77%	17.07[b]	53%	47%	.27	5.71[a]

Note. TC: Target cause, Assoc: Association between choice and condition.
[a] Significant at .05 level; [b] Significant at .01 level.

were identical to the TAs of Experiment 1. Sticking to the above example, participants had to choose between *Louis brought a turkey sub to the club* (TC1) and *Louis bought a soccer ball for the orphanage* (TC2) as the more likely explanation for Louis' mother's later disagreement. Both in the analogy and in the target-only condition, participants received the trials and the answer options in counterbalanced order.

4.2. Results and Discussion

Table 3 shows the percentages of participants that chose TC1 and TC2 for the analogy and target-only groups, together with their Chi square statistics.

Participants in the analogy group preferred TC2 to TC1 in five of our sets of materials, with no trend in the remaining set. Participants' choices thus departed from those that SME and LISA would favor. The preference for TC2 in the analogy group cannot be attributed to a higher intrinsic plausibility of TC2s as the causes of the target effects, since the target-only group showed a preference for the TC1s or no preference at all, yielding an association between TC choice and condition in five of the six sets of materials (see Table 3). Data thus replicate results obtained in Experiment 1, but this time using more typical analogies (i.e., tasks demanding the comparison of systems of relations), and a more ecologically valid task (i.e., identifying the cause of a given outcome).

5. General Discussion

A crucial component of analogical reasoning consists of assessing to what extent two situations should be considered analogous. The present study set forth to determine if the semantic assumptions postulated by the structure mapping theory (as implemented in SME; Falkenhainer et al., 1989; Forbus et al., 1994) and the multiconstraint theory (as implemented in LISA, Hummel & Holyoak, 1997, 2003), are adequate to account for how humans judge the quality of an analogy. Despite a number of differences maintained by these theories in other respects, they share the idea that, all other things being equal, two situations will be regarded as more analogous whenever the elements that conform one of the situations are similar to the corresponding elements in the other, as determined by the mapping process.

   As was thoroughly analyzed, if confronted with a BA stating that *John rented a car* followed by two alternative TAs, one stating that *Peter borrowed a van* (TA1) and another stating that *Richard broke a glass* (TA2), all of the above algorithms would choose TA1 as more analogous to the BA. While LISA would pick TA1 in response to the similarities maintained by matched relations and objects, SME would base its evaluation solely on the similarity between relations—a criterion that has received some empirical support (e.g., Gentner et al., 1993; Gentner & Kurtz, 2006). In this example, the similarity between BA and TA1 is almost completely explainable in terms of the preexisting similarity between the paired relations *rent* and *borrow* as well as between the concepts *car* and *van*—a resemblance that we have termed *element similarity*. In this sense, the criteria followed by the above models seems to adequately capture the intuition that two situations should be considered analogous to the extent that one can find a reasonably informative description capable of encompassing both situations, being the concatenation of relatively immediate superordinates a reasonably promising means of finding such a description. As the above example illustrates, simply combining the superordinates of the paired elements of BA and TA1 (i.e., *A person acquires a vehicle temporarily*) in fact affords a more informative description than would be obtained by combining the common superordinates of BA and TA2 (i.e., *A person exerts an action on*

an object).

We contended that an important limitation of the standard approach to semantics arises when it has to explain the assessment of the quality of analogies in which the interrelation between propositional elements in each analog invites interpretations and comparisons that go beyond the schema that would arise via assembling the superordinates of the paired elements. For instance, the sentence *John cut Peter's hair* invites an interpretation (an intervention to improve his look) very different from that of *John tore Peter's hair out* (an aggression). While the described algorithms would consider these two facts as analogous, people might not see them that way. In cases like this, simply assembling the superordinates of the paired elements can be misleading.

The failure of programs such as SME or LISA to reject comparisons which (as we conjectured) people would readily consider as non-analogous is only one half of the story. By the same token, the above programs would fail to capture what we have called overall similarity–a type of analogical relatedness that comes out of comparing the meaning of whole propositions, and that cannot be derived from similarities between paired propositional elements. Programs would fail to acknowledge, for instance, the similarity between *John cut Peter's hair* and *John sprayed perfume over Susan's hair* (interventions to improve his look)–a resemblance that, as we predicted, people would readily perceive.

Experiment 1 demonstrated that when deciding which of two TAs was more analogous to a BA, people pass over element similarities and seem to follow descriptors that cannot be derived from this type of similarities. Experiment 2 replicated these results embedding the first-order propositions of Experiment 1 within systems of relations, and replacing the task of evaluating the quality of an analogy between two simple facts with a more natural task that consisted of identifying the most likely cause of a target situation. Unlike SME and LISA, which would be biased towards element similarities, our participants proved to follow overall similarities.

Even though our results are silent as to the exact mechanisms underlying participants' sensibility to overall similarities during the evaluation of analogies, we will share some intuitions about their nature. It seems likely that participants respond to the task of assessing the quality of an analogy

by trying to assign the facts being compared to a schema-governed category of the type postulated by authors like Markman and Stilwell (2000; see also Goldwater, Markman & Stilwell, 2011). Instead of sharing a set of probabilistic features and feature correlations, members of schema-governed categories such as *assassination* share a structure like KILL (murder, means, victim), which can be instantiated by many apparently different exemplars, such as *Fred thrust a knife into Gina's heart*, *Mary had Bob drink poison*, or *The surgeon disconnected the patient's oxygen supply* (Gentner & Kurtz, 2005). The inclusion of an exemplar into this type of categories does not require that it maintained element similarities (be it of relations or objects) with other exemplars of the category—a feature that could potentially explain why two facts can be considered analogous despite the absence of element similarities. Consider the following analogs:

> BA: Delores hung garlic on the door
> TA: Mary brought a rabbit leg to the stadium

If people were to decide whether these situations are analogous via searching for superordinates for the pairs *hang* and *bring*, *garlic* and *rabbit leg*, and *door* and *stadium*, they would only find very abstract ones, giving place to a vapid description like "Someone takes an object to a place", probably leading participants to decide that the compared facts are not analogous. In cases like these, where traditional rerepresentational mechanisms such as minimal ascension or semantic decomposition fail to reveal an interesting identity between the compared situations, people may search for a schema-governed category (in this case, *superstitious behavior*) for which the to-be-compared situations could be considered instances. Given that the BA and the TA constitute relatively typical exemplars of superstitious behavior, it is likely that each analog can elicit such category on its own, easing the evaluation task considerably. However, it is possible that in some cases one of the analogs (e.g., the BA) constitutes a typical exemplar of a schema-governed category but the other one (the TA) does not, admitting the application of more accessible alternative categories. If the typical BA promotes a relatively improbable categorization of the TA, it could be taken as a case of recategorization. Consider, for example, the

following analogs (from Set 3 of our materials):

> BA: Delores hung garlic on the door
> TA: Mary lighted a candle in the basement

In cases like these, people are likely to categorize the BA as an exemplar of superstitious behavior (since it is a typical example of that schema-governed category), and then evaluate if the TA could be considered an instance of such category. This kind of recategorization is likely to occur for schema-governed categories, since the exemplars of these categories usually receive many and diverse categorizations (some of them not mutually exclusive), as compared to exemplars of entity categories (Gentner & Kurtz, 2005) (e.g., *Mary lighted the candle in the basement* could be categorized as an act of illumination, an attempt to improve the smell of the basement, etc.). If the BA represents a typical exemplar of a schema-governed category, it may favor the application of such category to the TA in order to reveal the similarity between the base and the target (see Oberholzer, Trench, & Minervino, 2011, for empirical evidence for this mechanism of analogical rerepresentation).

Finally, it could eventually happen that neither of the analogs is a typical case of a common schema-governed category:

> BA: The coach wore his red jacket during the match
> TA: Mary lighted a candle in the basement

In cases like these, as none of the analogs naturally elicits the superstitious behavior category on its own, it seems sensible to conjecture that the description of both analogs as instances of such category would result from searching for a schema-governed category capable of encompassing both situations, despite not being likely to be activated by either analog on its own.

On occasions, the schema-governed category finally applied to the base and target facts may not consist of a stable and lexicalized concept within the cognitive system, as evidenced by the difficulty implied in conceptualizing and verbalizing the reasons for accepting a given

comparison as analogical. Consider, for example, the following analogs:

> BA: Delores hung garlic on the door prior to the match
> TA: Mary prayed a lot prior to the match

While it seems easy to consider the BA a case of superstitious behavior and the TA as an instance of religious practice, it is not easy to say which is the shared schema-governed category (one way of describing it could be "two supernatural modes of preventing the occurrence of a non-desirable sport outcome").

A possible response to our critique of the standard model's account of how analogies are evaluated could argue that the analogical machinery was not meant to deal with the activity of comprehending the analogs, but rather to start operating once each of the analogs has been fully comprehended. In this vein, the categorization of situations like *Delores hung garlic on the door* and *Mary brought a rabbit leg to the stadium* as cases of *superstition* constitute preconditions for analogical reasoning to take place, and as such does not form part of the analogical process. Given that the comparison of the above situations takes as input representations already cast as HAVE [Delores, SUPERSTITIOUS (behavior)] and HAVE [Mary, SUPERSTITIOUS (behavior)], computer models could trivially account for participant's acceptance of their analogical relatedness in terms of similarities between corresponding propositional elements.

To the best of our understanding, the above objection faces a serious problem. A quick glimpse at our stimuli shows that, in contrast to this last example, in most sets at least one of the analogs cannot be considered a typical exemplar of the schema-governed category that is required to understand the analogy. For the objection to hold in these cases, the analogist would need to have represented the atypical analog as a member of a shared schema-governed category *before* the analogical comparison begins. As we have pointed out, *Mary lighted the candle in the basement* is most likely categorized as an act of illumination, an attempt to improve the smell of the basement, etc. The probability of having conveniently categorized such fact as an act of superstition at the very outset of the analogical process is rather low. The overestimation of this probability

could be an expression of 20-20 hindsight, which implies disregarding the interaction between mapping and representation-building in analogical thinking (see, e.g., Dietrich, 2000; Hofstadter & the Fluid Analogies Research Group, 1995). If we parsimoniously assume that most people interpret the TA as an attempt to illuminate the basement, the analogical process would then start off with representations like:

BA: HAVE [(Delores, SUPERSTITIOUS (behavior)]
TA: ILLUMINATE [(Delores, basement, candle)]

Given that the rerepresentational mechanisms incorporated by the standard models of analogical reasoning basically consist of finding the least common superordinate concepts of the to-be-paired elements, it seems obvious than none of these mechanisms can give rise to a meaningful description encompassing both analogs. Furthermore, since the above interpretations of the BA and the TA are not formally equivalent, such mechanisms could not even be applied. That said, the recategorization of the TA as a case of *superstitious behavior* is clearly the result of rerepresentational mechanisms operating at the level of entire propositions, and not at the level of propositional elements. As was already suggested, they are mechanisms that start off with an interpretation of the entire base proposition (or system of relations), and then proceed to reinterpret the TA in terms of the BA. This reply to the objection can of course be extended to those cases where an analogy promotes a rerepresentation of both analogs. Leaving behind a heated debate of the 90s, the idea that in most occasions representation-building and mapping run in parallel is now widely accepted (cf., e.g., Gentner & Kurtz, 2006; Kurtz, 2005; Kokinov, Vankov & Bliznashi, 2009), a state of affairs that should lead to discuss how this interaction might be approached by theories and computational models of analogical reasoning.

Regardless of the exact type of representations taken as input by the analogical engine, the question remains as to whether representing the analogs as instances of a shared schema-governed category is sufficient for evaluating the quality of analogies. It could be the case that beyond assigning the analogs to a common category, people still retain a rather

concrete and specific representation of each of the analogs, which can be quite useful in determining to what extent two analogous situations should be considered equivalent. Upon completing the experimental session of Experiment 1, we asked several participants to justify why they had chosen TA2 as a better match to the BA. Participants frequently evoked schema-governed categories quite similar to those we had in mind when constructing our materials, thus providing anecdotal support for our intuitions about the mechanisms involved in judging the quality of an analogical comparison. More interestingly, we found that despite having chosen TA2 as more analogous than TA1, participants were far from considering TA2 as perfectly analogous to the BA. For example, after comparing *The dictator had his servant try his meal* with *The dictator had his double attend the parade*, one participant argued that despite being "two cases of avoiding a potentially dangerous situation", they were quite different in the sense that "whereas the former case can undoubtedly be considered an extreme precaution and an abuse of authority, the latter request is a common activity of being a double". We conjecture that the process of evaluating the quality of an analogy operates at two levels: an upper level, at which the analogist decides whether the compared situations can be assigned to a schema-governed category, and a lower level, at which he or she assesses to what extent they can be considered similar instances within such category, probably by analyzing alignable differences (Gentner & Markman, 1994) along relevant dimensions of the category. In order to mimic human performance during tasks akin to those employed in our experiments, it seems that computer models would need to compare two analogous situations in terms of whether they belong to a common schema-governed category, but neither disregarding the more concrete level of representation nor losing track of the correspondences between both levels.

   Even though our intuitions about the mechanisms underlying how people evaluate the quality of analogies surely await a great deal of theoretical refinement and empirical testing, we believe that they open up interesting perspectives for the development of theoretical and computational models of analogical mapping and evaluation. An important start could consist in incorporating schema-governed categories to the knowledge base of the programs, and taking hand of this information when convenient. However,

in those cases where the shared schema-governed categories need to be constructed on the fly, the systems will need to handle large amounts of world knowledge in a creative way, a challenge that doesn´t seem to be attainable in the short run.

References

Bassok, M., & Medin, D. 1997. Birds of a feather flock together: Similarity judgments with semantically rich stimuli. *Journal of Memory and Language* 36, 311–336.

Dietrich, E. 2000. Analogy and conceptual change, or you can't step into the same mind twice. In E. Dietrich & A. Markman (Eds.), *Cognitive Dynamics: Conceptual Change In Humans And Machines*. Mahwah, NJ: Lawrence Erlbaum Associates, 265–294.

Falkenhainer, B. 1990. Analogical interpretation in context. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ, 69-76.

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

Forbus, K. D., Ferguson, R. W., & Gentner, D. 1994. Incremental structure-mapping. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, Atlanta, Georgia, 313-318.

French, R. M. 2002. The Computational Modeling of Analogy-Making. *Trends in Cognitive Sciences* 5, 200-205.

Gentner, D. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7, 155-170.

Gentner, D., & France, I. M. 1988. The verb mutability effect: Studies of the combinatorial semantics of nouns and verbs. In S. L. Small, G. W. Cottrell, & M. K. Tanenhaus (Eds.), *Lexical Ambiguity Resolution: Perspectives From Psycholinguistics, Neuropsychology, And Artificial Intelligence.* San Mateo, CA: Kaufmann, 343-382.

Gentner, D., & Forbus, K. 2011. Computational models of analogy. *WIREs Cognitive Science* 2, 266-276.

Gentner, D., & Kurtz, K. J. 2005. Learning and using relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. W. Wolff (Eds.), *Categorization Inside and Outside the Laboratory,* Vol. 43, 151–175, Washington, DC: APA.

Gentner, D., & Kurtz, K. 2006. Relations, objects, and the composition of analogies. *Cognitive Science* 30, 609-642.

Gentner, D., Holyoak, K. J., & Kokinov, B. N. 2001. *The Analogical Mind: Perspectives from Cognitive Science*. Cambridge, MA: MIT Press.

Gentner, D., & Markman, A. B. 1994. Structural alignment in comparison: No difference without similarity. *Psychological Science* 5, 152-158.

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

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

Gentner, D., Rattermann, M. J., & Forbus, K. D. 1993. The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology* 25, 431-467.

Goldwater, M. B., Markman, A. B, & Stilwell, C. H. 2011. The empirical case for role-governed categories. *Cognition* 118, 359-376.

Hofstadter, D. & the Fluid Analogies Research Group. 1995. *Fluid Concepts And Creative Analogies*. New York: Basic Books.

Holyoak, K. J. 1984. Analogical thinking and human intelligence. In R. J. Sternberg (Ed.) *Advances in the Psychology of Human Intelligence*, Vol. 2, 199-230. Hillsdale, N.J.: Erlbaum.

Holyoak, K. J., & Thagard, P. R. 1989. Analogical mapping by constraint satisfaction. *Cognitive Science* 13, 295-355.

Holyoak, K. J., & Thagard, P. R. 1995. *Mental Leaps: Analogy in Creative Thought*. Cambridge, MA: MIT Press.

Hummel, J. E., & Holyoak, K. J. 1997. Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review* 104, 427-466.

Hummel, J. E., & Holyoak, K. J. 2003. A symbolic-connectionist theory of relational inference and generalization. *Psychological Review* 110, 220-264.

Keane, M., Hackett, D., & Davenport, J. 2001. Similarity processing depends on the similarities present. In J. D. Moore & K. Stenning (Eds.), *Proceedings of Twenty-Third Annual Conference of the Cognitive Science Society*. Mahwah NJ, 477-482.

Kersten, A. W., & Earles, J. L. 2004. Semantic context influences memory for verbs more than memory for nouns. *Memory and Cognition* 32, 198–211.

Kokinov, B., Vankov, I., & Bliznashki, S. (2009). How Analogy Could Force Re-representation of the Target and Inhibition of the Alternative Interpretation. In B. Kokinov, K. Holyoak & D. Gentner (Eds.) *New Frontiers in Analogy Research*. Sofia: NBU Press, 269-278.

Larkey, L. B., & Love, B. C. 2003. CAB: Connectionist analogy builder. *Cognitive Science* 27, 781-794.

Markman, A., & Stilwell, C. 2001. Role-governed categories. *Journal of Experimental & Theoretical Artificial Intelligence* 4, 329–358.

Minervino, R. A., Oberholzer, N., & Trench, M. 2008. Similarity between propositional elements does not always determine judgments of analogical relatedness. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society, 91-96.

Oberholzer, N., Trench, M., & Minervino, R. A. (2011). When lighting a candle becomes a superstition: Analogical recategorization through the application of relational categories. *33rd Annual Meeting of the Cognitive Science Society*. Boston, Massachusetts, 568–573.

Yan, J., Forbus, K., & Gentner, D. 2003. A theory of rerepresentation in analogical matching. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*. Mahwah, NJ, 1265–1270.