



The Role of Surface Similarity in Analogical Retrieval: Bridging the Gap Between the Naturalistic and the Experimental Traditions

Máximo Trench, Ricardo A. Minervino

Department of Psychology, University of Comahue

Received 6 March 2014; received in revised form 23 June 2014; accepted 16 July 2014

Abstract

Blanchette and Dunbar (2000) have claimed that when participants are allowed to draw on their own source analogs in the service of analogical argumentation, retrieval is less constrained by surface similarity than traditional experiments suggest. In two studies, we adapted this production paradigm to control for the potentially distorting effects of analogy fabrication and uneven availability of close and distant sources in memory. Experiment 1 assessed whether participants were reminded of central episodes from popular movies while generating analogies for superficially similar versus superficially dissimilar target situations. In Experiment 2, we modified this procedure to assess the retrieval of autobiographic memories, more familiar to participants than fictional episodes. Both studies revealed a strong effect of surface similarity on the retrieval of participants' own sources, thus suggesting that the superficial bias typically observed in experimental studies—and simulated by most computational models—does not originate in a lack of ecological validity.

Keywords: Analogy; Analogical retrieval; Superficial similarity; Naturalistic settings

1. Introduction

Across activities as diverse as problem solving, instruction, or argumentation, analogical reasoning allows knowledge transfer from a known situation (the *base analog*, or *source analog*) to a less known situation (the *target analog*). This transfer process is based on the recognition that the entities of the base and target analog are linked by systems of relations that can be considered formally and semantically identical at a

Correspondence should be sent to Máximo Trench, Department of Psychology, University of Comahue, Quintral 1250, CP (8400), S. C. de Bariloche, Rio Negro, Argentina. E-mail: maxtrench@gmail.com

Results from Experiment 1 were presented at the 33rd Annual Conference of the Cognitive Science Society, Boston, MA, 2011.

non-trivial level of abstraction (Gentner, 1983; Gentner & Markman, 2006; Holyoak & Thagard, 1995). As an example, the Greek philosopher Cornelius Castoriadis compared the consumption of natural resources by capitalist societies with Hansel and Gretel's tale by the Brothers Grimm: Just as Hansel and Gretel were eating the chocolate walls of their house without knowing that they were destroying their house, capitalist nations are devastating the forests without realizing that they are disturbing climate regulation. The analogy rests in the fact that in both cases, a valuable thing is being carelessly damaged. Whereas in *distant analogies* the corresponding objects and relations usually do not maintain semantic similarity (e.g., Hansel ↔ Nations; eat ↔ devastate, chocolate walls ↔ forests), in *near analogies* the corresponding elements are semantically similar. As an example, the environmental movie "Home" compared the devastation of forests by current governments to the exhaustion of palm trees by the original inhabitants of Easter Island. As opposed to the Hansel and Gretel analogy, in this case structural similarity comes together with a higher degree of semantic similarity (e.g., inhabitants ↔ governments, exhaust ↔ devastate, palm trees ↔ forests), thus allowing a subsuming schema that is less abstract than in the previous case. These lower level semantic similarities have been termed *superficial similarities*, under the consideration that they are negligible when more abstract identities are taken into account. In the present study, we will refer to these types of comparisons as *superficially dissimilar* and *superficially similar* analogies, respectively.

Among researchers of analogical reasoning, there is ample consensus that the process of understanding an analogy is not heavily dependent on superficial similarities (Gentner, Rattermann, & Forbus, 1993; see Holyoak, Novick, & Melz, 1994 for a review). As Castoriadis's analogy illustrates, the lack of semantic similarity at the level of matched objects and relations does not necessarily complicate the process of finding the right correspondences between the analogs, since a reasonably informative description encompassing both situations can be easily derived. In contrast to the consensus about the (minor) role played by superficial similarities during mapping, the weight of superficial similarities during the retrieval of base analogs from long-term memory (LTM) is still subject to some debate. Whereas a wealth of experimental studies showed that superficial similarity represents a crucial precondition for retrieval (e.g., Gentner et al., 1993; Keane, 1987; Ross, 1989), more recent naturalistic studies allegedly demonstrate that the retrieval of participants' own sources during real-world tasks like persuasive argumentation does not necessitate superficial similarities (e.g., Blanchette & Dunbar, 2000, 2001). This empirical inconsistency between the results of the experimental and the naturalistic approaches to analogical retrieval—termed "The Analogical Paradox" by Dunbar (2001)—has not been resolved. The present study aims to shed light on the reasons for this inconsistency by means of a hybrid method that preserves the strengths of these two empirical traditions. Before moving to our own study, we discuss the procedures followed by the experimental and the naturalistic traditions, and briefly describe how extant computational models have attempted to account for the observed behavioral patterns of analogical retrieval.

1.1. *The experimental tradition*

As in most studies of memory, the standard paradigm for investigating analogical retrieval comprises two distinct phases, an encoding phase and a retrieval phase. During encoding, participants receive the base analogs (thus the expression “reception paradigm” coined by Dunbar, 2001), usually interleaved by several distracters. During the retrieval phase, participants are presented with the target analog embedded in target tasks for which the retrieval of the critical base analog becomes useful, and experimenters assess the extent to which the processing of the target triggers the retrieval of the base analog. Materials, tasks, and contextual/temporal separations between both phases vary considerably within this tradition. In studies of problem solving, the encoding of the base problem and its solution is typically followed by a contextual separation in order to guarantee that later access to learned solutions during the retrieval phase was not based on an episodic association between the phases. Results from these studies show that the retrieval of superficially similar sources is between two and four times more frequent than the retrieval of superficially dissimilar ones (e.g., Keane, 1987; Ross, 1989). As opposed to studies of problem solving, studies on the retrieval of stories rarely impose a contextual separation between the phases, since the salience of the critical sources is reduced by means of embedding them among a large number of distracter stories. During the retrieval phase, participants are handed each of the target analogs and are tasked with indicating which of the stories read during the previous session this new story reminds them of. Just as within the problem-solving literature, studies of story-reminding typically show that sources bearing superficial similarity with the target are much more likely to be retrieved than the sources that do not maintain such similarities (e.g., Catrambone, 2002; Gentner et al., 1993). The results from the problem-solving and the story-reminding tradition led researchers to conclude that superficial similarity represents a crucial precondition for analogical retrieval, with accounts for this superficial bias ranging from the evolutionary to the computational.

In terms of adaptation, failing to retrieve most of the distant matches stored in LTM is believed to represent no big loss, since superficial similarities tend to correlate with deeper commonalities in the natural world (the “kind world hypothesis,” Gentner, 1989). In the words of Dedre Gentner (1989, p. 267), “By and large, if something looks and roars like a tiger, it probably *is* a tiger.” Besides this ecological interpretation, computational modelers of analogical reasoning have claimed that the computational cost that would be implied in performing a structural match between the target and every potential source in LTM lacks psychological plausibility (Gentner & Forbus, 1991; Thagard, Holyoak, Nelson, & Gochfeld, 1990).

Under this last consideration, proponents of the structure mapping theory (Gentner, 1983; Gentner & Markman, 1997) developed MAC/FAC (Forbus, Gentner, & Law, 1994), an algorithm designed to simulate behavioral patterns of analogical retrieval through psychologically realistic computations. MAC/FAC, for *Many Are Called, Few Are Chosen*, divides retrieval into two distinct phases: MAC, a fast superficial filter, and FAC, a formal matcher. The MAC phase works by comparing the content vector for the

target with those of every representation stored in LTM, with each content vector being generated by assigning a position in an ordered series to all concepts in LTM, and counting how many times each of these concepts appears in each situation stored in LTM. Upon taking the vector products between the content vector of the target and the vectors of all situations in LTM, the MAC stage submits the winning situations (most of them superficially similar to the target) to the FAC stage. For each of these potential sources, FAC starts by creating all possible local mappings between elements of the same formal type (i.e., objects with objects, n-place relations with n-place relations, etc.), with the added restriction that mapped relations must have identical meaning. The program then incrementally coalesces local matches into global mappings that satisfy the constraints of *parallel connectivity* (if two predicates are mapped, their arguments must also be mapped) and *one-to-one mapping* (elements must map to at most one element in the other analog). Finally, FAC scores the quality of global mappings as a function of their size, their depth, and the semantic similarity of their corresponding objects. This last criterion amplifies MAC's bias toward base analogs bearing superficial similarity with the target.

LISA (*Learning and Inference with Schemas and Analogies*; Hummel & Holyoak, 1997) is the latest matcher developed by proponents of the multiconstraint theory of analogy (Holyoak & Thagard, 1989, 1995). Its architecture aims at encompassing retrieval, mapping, inference, and schema abstraction by a unified set of core processes more neurally plausible than in earlier attempts (e.g. ARCS; Thagard et al., 1990). LISA's architecture is a system for representing dynamic role-filler bindings in working memory (WM) and encoding them in LTM for later retrieval. When a proposition unit (P) like *John loves Mary* gets activated, it propagates top-down activation to subproposition units (SPs) that represent 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 (e.g., *John* and *lover*) which fire in synchrony with each other and out of synchrony with the units of the complementary SP (i.e. *Mary* and *beloved*). Case roles and their fillers—which represent the lowest level in the structural hierarchy—in turn activate a collection of semantic units representing their meaning. 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 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. Syntactic constraints are enforced by sets of excitatory and inhibitory links. Within a base analog, units of different hierarchy are linked by symmetric excitatory connections, whereas units of the same level share symmetric inhibitory links. This way, when predicate and object units in a base analog respond to patterns of activation in WM, they activate SP and P units above them, all of which inhibit other units of the same type, enforcing the one-to-one mapping constraint. Once a P unit in the target has activated a corresponding P unit the base

analog, the constraint of parallel connectivity is enforced by top-down activation of the structure units below them. As in MAC/FAC, LISA's reliance on semantic similarities between the sources and the target leads to a majority of superficial reminders.

In sum, besides the contrasting representational and computational assumptions incorporated in the above models (e.g., MAC/FAC uses serial operations on symbolic representations, while LISA uses connectionist computation on distributed representations), both algorithms were engineered to account for the observed behavioral patterns obtained with the classical two-phase reception paradigm by means of granting a crucial weight to superficial similarities.

1.2. *The naturalistic tradition*

With the turn of the century, a number of observational studies began documenting the use of analogies by professionals of different disciplines as they worked within their areas of expertise. While the analogies produced by molecular biologists (Dunbar, 1997) and psychologists (Saner & Schunn, 1999) tended to be highly constrained by superficial similarity, it became apparent that the goals of the analogizer might have an effect on the semantic distance of the generated analogies. More recent observational studies demonstrated that under specific pragmatics such as communicating ideas to others, the use of distant analogies can be as frequent as the use of superficially similar analogies.

In a seminal study on the use of analogies in political discourse, Blanchette and Dunbar (2001) analyzed more than 400 articles that appeared in three important newspapers from Montreal prior to the 1995 referendum on the independence of Quebec. The authors found that three quarters of the more than 200 analogies that appeared in the media involved situations outside the domain of politics, such as agriculture, family, sports, magic, and religion—a result that was interpreted as indicating that naturalistic analogy generation is less constrained by superficial similarity than in the experimental tradition.

Richland, Holyoak, and Stigler (2004) set forth to investigate the spontaneous use of analogies in eighth-grade mathematics classrooms. Among other results, they observed that the degree of surface similarity of the sources used by the teachers was highly dependent on the type of knowledge they were trying to impart. While the teaching of procedures was based mainly on superficially similar analogies, transmission of conceptual knowledge tended to be based on superficially dissimilar or even non-mathematical analogies. The fact that, overall, 76% of the analogies had minimal perceptual similarity between source and target objects was taken by the authors to suggest that “in the classroom setting, teachers are more successful than typical laboratory participants at developing structural analogies” (Richland et al., 2004, p. 49).

More recently, Christensen and Schunn (2007) carried out a systematic observation of the analogies generated by professional design engineers as they created new products in the domain of medical plastics. During the weekly product development meetings, which lasted between 30 min and 2 h, common activities included brainstorming, developing concepts, solving design problems, planning data collection, evaluating mock-ups and prototypes, among many others. The transcripts of these meetings revealed a mean of

11.3 analogies per hour of verbal data. Superficially dissimilar analogies were almost as frequent as superficially similar analogies when the goal was to solve a problem, and twice as frequent when the goal was to communicate or explain ideas to other members of the group.

According to various authors (e.g., Dunbar, 2001; Hofstadter & Sander, 2013; O’Keefe & Costello, 2008), the results of naturalistic studies call into question the psychological validity of laboratory experiments failing to obtain distant retrievals, as well as the accuracy of the psychological theories and computational models developed after such behavioral results. In order to simulate the pattern of distant analogizing observed in naturalistic studies, O’Keefe and Costello (2008) developed a computational model that relies exclusively on similarities between structurally equivalent systems of relations. Based on the assertion that “contemporary theories of memory do not posit that memory is composed of distinct domains [the base analogs] as assumed in MAC/FAC” (O’Keefe & Costello, 2007, p. 5), the authors adopt a model of LTM that consists of a list of all instances of any given relation (regardless of the base analogs in which they take part), which are in turn linked to the relations that were adjacent to them in the episodes that were encoded in LTM. As there is no explicit separation between the base analogs stored in memory, the model’s current version can only deal with *connected representations*—that is, base and target analogs in which every relation is directly or indirectly linked to every other relation by means of one or more propositional links. Fed with this type of representations, O’Keefe and Costello’s (2008) relational adjacency model performs retrieval and mapping in an intertwined fashion according to the following procedure. Given a target analog represented in higher order relational form, the system starts by choosing an arbitrary relation in the target (e.g., *relation #1*) and selecting all instances of such relation in memory. Next, the system shifts its attention back to the probe, and after identifying any relation (or relations) that are adjacent to relation #1 in the probe (e.g., relations #2 and #3) it trims the initial cohort of instances of relation #1 in LTM, preserving only those that are adjacent to both relation #1 and #2. Using the same logic, the system continues checking to see if the relations adjacent to each instance of a relation in the probe are adjacent to the last level added to the partial matches in memory. When all the vertices in the probe have been visited, the search is complete and any complete matches will have been found.

In its current state of development, O’Keefe and Costello’s algorithm can only retrieve base analogs being completely isomorphic to the target. As a consequence of this limitation, the system is incapable of simulating any kind of analogical transfer, since transfer requires that elements in the base analog lack their counterparts in the initial representation of the target. Even though this state of affairs is certainly consistent with Blanchette and Dunbar’s (2001) and Richland et al.’s (2004) finding that in natural settings, analogizers use semantically distant sources to communicate ideas to others (a situation where the reasoner starts-off with a complete representation of the target idea she wants to communicate), it falls short of accounting for Christensen and Schunn’s (2007) more recent finding that distant analogies are also frequent in situations in which an impoverished

target analog gets illuminated by a base analog whose representation is comparatively more complete (e.g., during creative problem solving).

To summarize, a rapidly growing body of naturalistic studies on the spontaneous use of analogy in real-world activities tends to show that, with few exceptions, reasoners can flexibly access superficially dissimilar sources from their background knowledge in a manner that is certainly not predicted on the basis of traditional experimental results (e.g., Gentner et al., 1993; Keane, 1987), and which cannot be simulated by dominant computational models of analogical retrieval like MAC/FAC or LISA.

1.3. *Bridging the gap between the naturalistic and the experimental traditions*

There is a well-known trade-off between the strengths and weaknesses of laboratory versus naturalistic studies: while experimental control sometimes comes at the expense of ecological validity, the richness of naturalistic observation usually comes at the expense of control. To overcome this conundrum, various authors have advocated for “cross-fertilizing” both approaches by means of going back-and-forth between experimental and naturalistic approaches (i.e., the *in vivo/in vitro approach*, Dunbar, 2001; Klahr & Li, 2005). Naturalistic observation can be useful for uncovering interesting phenomena and relations between variables that would have remained undetected by means of purely experimental hypothesis testing. In turn, reproducing these findings within laboratory settings can be useful for scrutinizing the observed phenomena under more controlled conditions, and for discriminating those correlations that are based on causal relations from those that are not.

According to Dunbar (2001), the use of an *in vivo/in vitro* methodology was instrumental in uncovering the reasons behind the disparate results from the experimental and the naturalistic traditions in the study of analogical retrieval. In a series of laboratory experiments carried out after having studied the use of analogy by journalists and politicians, Blanchette and Dunbar (2000) asked groups of college students (Experiment 1) and individual students (Experiment 2) to generate persuasive analogies for another real-world political topic: the zero-deficit strategy. After providing participants with an explanation about the mechanisms by which public debts typically increase, participants in the pro zero-deficit condition had to pretend they were hired by a non-profit organization to generate analogies that could be used to persuade the population of supporting massive cuts to spending in education, security, and social programs on the grounds that future cuts would otherwise be more dramatic. In contrast, participants in the anti-zero-deficit condition were asked to propose analogies to persuade the public that basic services cannot be discontinued, and that other ways of dealing with the deficit should be envisioned. Regardless of the condition, this *production paradigm* (as Blanchette and Dunbar termed it) elicited an overwhelming majority of base analogs pertaining to domains different from economy and politics, thus reproducing *in vitro* the previous naturalistic observation that journalists can easily access long-distance sources while generating analogies to real-world target situations. In order to assess whether the success of their subjects in generating superficially dissimilar analogies was due to some degree of familiarity with

the zero-deficit topic—this issue had received extensive media coverage in the months prior to the study—Blanchette and Dunbar (2000, Experiment 3) resorted to a traditional reminding task, in which the superficial (or structural) base analogs to be read during the encoding phase were chosen among those proposed by participants in Experiments 1 and 2. In contrast to the type of analogies proposed under a production paradigm (Experiments 1 and 2), participants in Experiment 3 retrieved mostly non-analogical matches bearing only superficial similarity with the target. These results were interpreted as demonstrating that the central feature behind the success of participants in retrieving distant analogs in Experiments 1 and 2 was not their degree of familiarity with the target topic, but rather the fact that they were allowed to draw on their own extraexperimental knowledge to cope with an ecologically valid task such as generating persuasive analogies in favor of certain line of action. Dunbar (2001) suggested that, as naturalistic settings and ecologically valid tasks can promote an abstract encoding of the base and target analogs, respectively, retrieval of sources is not heavily constrained by superficial similarity. More recently, Markman, Taylor, and Gentner (2007) have demonstrated that an auditory presentation of both the base and the target analogs leads to better recall of purely relational matches than a written presentation. Based on these results, the authors conjectured that one of the factors underlying the apparent ease of relational retrieval in naturalistic settings could be the fact that in natural settings, as opposed to typical experimental studies, the analogs tend to be presented in auditory, rather than in written format.

As we will argue, even though the production paradigm implemented by Blanchette and Dunbar (2000) can potentially exert a higher degree of control than purely observational approaches,¹ it falls short of demonstrating a negligible role of superficial similarities during naturalistic retrieval. Compared to the traditional two-phase procedure followed by most experimental studies of analogical retrieval, the production paradigm bears two important methodological shortcomings. The first limitation resides in not implementing any means of distinguishing true instances of analogical retrieval from instances of analogy fabrication (i.e., ad hoc invention of base analogs). For example, once the reasoner has understood the abstract structure of a target situation, he or she can trivially generate novel exemplars from semantically distant domains (e.g., infections of HIV/typhus/choose your favorite infection/that were not treated in time, with the consequence that later treatment was more difficult that would have been otherwise). Second, even if a means of distinguishing true retrievals from cases of analogy fabrication were implemented, yet another serious insufficiency of the production paradigm concerns the unknown proportions of close versus distant base analogs potentially available in LTM. As a hypothetical example, consider a participant retrieving six near and six distant base analogs during the analogy generation task. In case this participant had, say, 10 near and 10 distant sources stored in LTM, the retrieval probabilities of near and distant base analogs could be considered superficially unconstrained (60% of each type). In sharp contrast, if he or she had 60 near sources and 10 distant sources available in LTM, the very same retrieval outcome would now indicate that the probability of retrieving base analogs is superficially biased (10% of superficially dissimilar vs. 60% of superficially similar sources). Therefore, a proper assessment of the effect of superficial similarity on

analogical retrieval requires knowing not only the quantities of superficially similar and superficially dissimilar base analogs that were successfully retrieved but also the number of instances of both types of base analogs that were potentially available for retrieval (i.e., retrieval probability = number of retrieved sources/number of available sources). As can be inferred from the above considerations, current attempts to explain out the divergence between experimental and naturalistic findings fall short of demonstrating that in naturalistic settings analogical retrieval is less constrained by superficial similarity than in the psychological laboratory.

In the spirit of the *in vivo/in vitro* approach advocated by several authors, we sought to shed light on the inconsistency between naturalistic and experimental studies through a hybrid procedure that retains the ecological validity of the production paradigm, without sacrificing the inferential power of the reception paradigm to assert causal relations.

Just as in the production paradigm, participants of the two experiments reported in the present study had to generate persuasive analogies in response to a highly realistic target situation—an activity that allegedly promotes an encoding of the targets that emphasizes their abstract structure. Second, participants were asked to draw on their own extraexperimental memories, which according to authors such as Dunbar (2001), O’Keefe and Costello (2008), or Hofstadter and Sander (2013), receive a more abstract encoding than the base analogs typically used in traditional two-phase experiments. Having retained these distinctive advantages of the production paradigm as originally implemented, our studies also preserved those features of the reception paradigm that afford controlling the potentially distorting effects of analogy fabrication and uneven availability of near versus distant base analogs in LTM. In order to control for the uneven availability of close versus distant sources in LTM, in Experiment 1 we calculated the probability of being reminded of specific episodes from popular movies (the sources) during the task of generating persuasive analogies for realistic situations that maintained different degrees of superficial similarity with such episodes (the targets). By means of restricting the data analysis to the retrieval (or not) of such specific base analogs, we were able to calculate (and therefore compare) the accessibility of close versus distant naturalistic sources in terms of quotients between the number of retrieved cases and the number of cases available for retrieval.² On the other hand, the improbability of inventing a situation identical to a culturally shared episode helps ensure that all analogies built upon those episodes were originated in the retrieval of their representations from memory.

2. Experiment 1

2.1. Method

2.1.1. Participants and design

An initial cohort of 372 students of Psychology at Universidad Nacional del Comahue, Argentina, volunteered to take part in the experiment. The final sample was determined after all participants in the initial cohort had completed all the phases of the

administration. This final sample consisted of 160 participants who knew the critical base analog, and who were also able to make the analogy between such base and the target (see the Procedure section for further details). The degree of superficial similarity between each base analog and its two corresponding targets (high or low) received between-subjects manipulation. The dependent variable was the retrieval of the sources during the analogy generation phase.

2.1.2. *Materials and procedure*

Base analogs consisted of central episodes from *Spiderman*, *Shrek*, *Jurassic Park*, and *The Secret in Their Eyes*. For example, in *Jurassic Park* a millionaire has cloned dinosaurs from the Jurassic Period out of fossil DNA taken from a mosquito. Despite receiving expert advice about the impossibility of exerting total control over biological phenomena, the millionaire insists on opening a park to exhibit the dinosaurs to the public. Finally, dinosaurs break the security system of the park and attack human beings. Superficially similar targets were generated replacing base objects and relations with similar ones. For instance, the superficially similar target for *Jurassic Park* stated that a businessman had replicated mammoths from the Pleistocene Era out of a frozen embryo found in a glacier. The target ended up stating that the businessman persevered with his idea of opening a zoo with mammoths on show. The participants' task consisted of dissuading the main character from pursuing the project, warning him that as animal behaviors are not completely manageable, mammoths could destroy the zoo cages, thus endangering people.

Superficially dissimilar target analogs were derived replacing base objects and relations with objects and relations less similar than in the near targets. Continuing with the *Jurassic Park* set, the target stated that an astrophysicist was imitating Martian storms out of digital images captured by a space probe. The target ended up stating that he was planning to let his colleagues enter the experimental zone in order to study these storms. Participants had to dissuade the main character from pursuing his plan on the grounds that, as extraterrestrial climatic phenomena are not well known, they could exert negative effects on his colleagues. In both conditions, the instructions enforced participants to generate analogies that could be used to dissuade the character from carrying out his intended action. The complete targets derived from the movie *The Secret in Their Eyes* are displayed in Table 1. As can be observed, both target analogs were equally isomorphic to the base situation.

Participants received a booklet with the materials and tasks. The first two pages consisted of instructions on the use of analogy in persuasion, together with two examples in which the sources were real stories (one of them superficially similar to its target and the other one dissimilar) and two examples in which the sources were fictional stories (one of them superficially similar to its target and the other one dissimilar). This way, we tried to avoid biasing memory search neither in favor of fictional versus real sources, nor in favor of superficially similar versus superficially dissimilar ones. After reading the instructional material for 7 min, participants were allotted 15 min to read the target and write down as many analogies as they could generate to dissuade the character of the

Table 1
A complete example of a natural base analog and its derived target analogs, Experiment 1

Base Analog: <i>The Secret in Their Eyes</i>	Superficially Similar Target Analog	Superficially Dissimilar Target Analog
Some time ago a murderer killed Ricardo's wife	Some time ago an assailant maimed Sonia's father	Some time ago Romania defrauded the Bulgarian Government
A US Marshal said to Ricardo that the murderer would be in jail for life	A State's Attorney told Sonia that the assailant would be in prison for 25 years	A UN diplomat assured Bulgaria that Romania would be kept under a trade embargo for 10 years
However, due to irregular procedures, the murderer was set free soon afterward	However, due to non-official mechanisms, the assailant was released soon afterward	However, due to complex negotiations the embargo was lifted soon afterward
Since then, Ricardo's only obsession has been ensuring that the murderer serves the original sentence	Since then, Sonia's sole preoccupation has been having the assailant fulfill the preestablished penalty	Since then, Bulgaria's main interest has been having Romania undergo the stipulated restriction
For years, Ricardo has locked the murderer up himself	Sonia is planning to shut the assailant up by herself	Bulgaria is planning to boycott Romania on its own
This has led Ricardo to give up important personal ambitions, such as finding a partner	You are a close friend of Sonia's, and you are concerned about her situation. You believe that pursuing her plan would lead Sonia to lose out on her individual aspirations such as building a family. You should evoke analogous situations or stories that you know, and that could be used to convince Sonia that executing her plan will entail giving up her own goals	You are a political consultant for Bulgaria, and you are concerned about its situation. You believe that pursuing its plan would lead Bulgaria to delay its own political programs, such as reforming the economy. You should evoke analogous situations or stories that you know, and that could be used to convince Bulgaria that executing such a plan will entail giving up its own goals

Notes. The base analog comprises the central episodes of the *The Secret in Their Eyes* that are relevant for establishing an analogy with the target analogs. The exact wording of the base analog is arbitrary, since the abstract representations stored in memory may not be specified at a lexical level. Italicized words in the base analog indicate objects (e.g., wife), object attributes (e.g., original), and relations (e.g., kill) that were replaced with either very similar concepts (superficially similar target) or with less similar concepts (superficially dissimilar target), also in italics.

story from carrying out his plan, warning him about a possible negative consequence of such a plan. Once this time had elapsed, participants had to answer a questionnaire aimed at detecting whether or not they had retrieved the critical source despite not having included it among their final proposals. To that end, they were asked whether they had been reminded of any movie during the analogy generation activity. In case they had, they were asked to indicate which movie or movies they were reminded of, and to state exactly which parts of such movie or movies they remembered at that time. Participants then answered a questionnaire aimed to assess whether they knew the specific facts about the source that were required to establish an analogy with the target. They were asked in the first place if they had seen the critical movie (those answering “no” finished the experiment right away). In case they had seen it, they continued to the next page where they had to answer 10 multiple-choice questions about its plot, having four options each. Finally, the last page of the booklet consisted of a task aimed at evaluating if participants were able to make the analogy between the movie and the target upon explicit request. The right column of a two-column table listed the six central actions of the target. Participants had to fill in the fields of the left column with the corresponding episodes of the movie. As the calculation of the retrieval probabilities should only take into account those cases in which the source was available for retrieval (see, e.g., Gentner et al., 1993), the final sample was limited to participants that got right 9 out of 10 questions of the availability questionnaire. In a similar vein, because retrieval trials should exclude the cases where a participant cannot make the analogy even when asked to do so, the analysis was limited to cases in which the participant got at least five out of six fields right in the final analogy-making task. In order to meet these controls, participants were run individually until completing 8 groups of 20 participants that, for the particular movie whose retrieval had been evaluated, demonstrated knowing the critical episodes of the movie that were analogous to the target, as well as being able to understand the analogical relation between the movie and the target. In four of these groups, participants received target situations that were superficially similar to the source they were expected to retrieve, and in the remaining four groups participants received targets that were superficially dissimilar from it. Table 2 summarizes the main phases of the administration, as well as their methodological justification.

2.2. *Results and discussion*

A base analog was scored as having been retrieved in the cases where the participant: (a) employed the base analog among the proposed analogies, or (b) reported having been reminded of the critical movie and its analogy-relevant facts despite not having included them among their proposals. Two independent judges, instructed in the six critical facts of each of the four movies, had to decide in which cases the participants included at least four of those facts in either their initial arguments or in their answer to the retrieval questionnaire. Judges agreed in 81% of the cases and solved the cases of disagreement by open discussion.

Table 2

Schema of the different phases of the procedure of Experiment 1, with their methodological purpose

Phases of the Experimental Procedure	Methodological Purpose
1. Participants receive a definition of analogy and an explanation of its utility for persuasion, illustrated with examples	Make participants aware that comparisons between analogous situations can be useful for persuasion
2. After reading a situation admitting two lines of action, participants were asked to support one of such actions by analogizing to known episodes. The situation was either a superficially similar or a superficially dissimilar analog of a popular movie seen by most participants	Involve participants in the ecologically valid task of analogical persuasion, and document the use (or not) of a specific base analog previously acquired in natural settings
3. Then they were asked whether they had been reminded of a movie during the prior task. In case they had, they were asked to specify what movie and what aspects of that movie they were reminded of during the prior task	Identify those cases where the critical source was in fact retrieved during the prior task, despite not being included among the analogical proposals
4. Participants answered questions about the particular movie that the experimenters had chosen as either the superficially similar or superficially dissimilar source for the target situation	Discard participants who did not demonstrate having a representation of the critical base analog in long-term memory
5. Participants were finally asked to analogize between the movie and the target scenario	Discard participants who were not capable of establishing the intended analogy upon explicit request

Superficially similar source analogs were retrieved in 70% of the trials (91% of the retrieved analogs were employed during the argumentation task). In contrast, superficially dissimilar sources were retrieved in only 15% of the cases (75% of the retrieved analogs were used for argumentation), thus demonstrating a strong effect of superficial similarities on the retrieval of the base analogs, $\chi^2(1, N = 160) = 47.29, p < 0.001$. This pattern of results held for each of the four base analogs that were employed. The plot of *The Secret in Their Eyes* was retrieved in 75% of the cases after a target with which it maintained superficial similarities, and in only 10% of the cases after a target without such similarities, $\chi^2(1, N = 40) = 14.73, p < 0.001$. In turn, *Shrek* was retrieved in 75% of the trials after a superficially similar target, and in 15% of the trials after a superficially dissimilar one, $\chi^2(1, N = 40) = 12.22, p < 0.001$. *Spiderman* was retrieved in 60% of the cases after a superficially similar target, and in 20% of the cases after a superficially dissimilar target, $\chi^2(1, N = 40) = 5.10, p < 0.05$. Finally, *Jurassic Park* was retrieved in 70% of the trials after a superficially similar target, and in 15% of the trials after a superficially dissimilar one, $\chi^2(1, N = 40) = 10.23, p < 0.001$.

This strong effect of superficial similarities on the retrieval of naturally acquired base analogs is consistent with a long experimental tradition using artificial stimuli (e.g., Keane, 1987; Gentner et al., 1993) and runs counter to the results obtained by Blanchette and Dunbar (2000) and simulated by O'Keefe and Costello's (2008) algorithm for superfi-

cially unconstrained retrieval. Results thus suggest that the alleged advantage of the production paradigm (and of naturalistic settings in general) in eliciting retrieval of superficially dissimilar sources does not originate in the fact that participants draw on their own memories during ecologically valid tasks such as persuasive argumentation, but rather in the effect of insufficiently controlled variables such as ad-hoc invention of analogous sources or uneven availability of superficially similar versus dissimilar base analogs.

A possible limitation of the materials used in Experiment 1 might result from the fact that the fictional stories used as natural sources, although probably better learned than the fictional stimuli employed in experimental studies (e.g., Gick & Holyoak, 1980), might have lacked the familiarity and meaningfulness that characterize non-fictional natural analogs. Hofstadter and Sander (2013) have recently argued that when dealing with real-world situations, the semantically distant source situations people retrieve spontaneously and effortlessly from their own memories are, in general, extremely familiar. According to the above authors, when making analogies, we all depend on knowledge that is rooted in our experiences over a lifetime, and this knowledge, which has been confirmed and reconfirmed over and over again, has also been generalized over time, allowing it to be carried over fluidly to all sorts of new situations. In their own words: "It is very rare that, in real life, we rely on an analogy to a situation with which we are barely familiar at all" (Hofstadter & Sander, 2013, p. 339).

3. Experiment 2

In the spirit of Hofstadter and Sander (2013), we adapted the procedure of Experiment 1 to assess the influence of superficial similarities on the retrieval of autobiographic memories, which in our intuition were more familiar and meaningful than the fictional stories employed in Experiment 1. Our strategy for assessing the retrievability of close versus distant autobiographic episodes comprised two phases. During the first phase, we asked participants to recall their own experiences conforming to several generic descriptions of events. In a second phase, temporally and contextually separated from the first, participants were confronted with hypothetical situations that were also cases of those same generic types of situations, with the instruction of using analogies to dissuade the main character of such situations from carrying out his intended action. Target situations were thus structurally equivalent to the base situations reported during the prior phase. Depending on the condition, however, target situations were superficially similar or superficially dissimilar from the instances reported by participants during the previous phase. A further indication to base their analogies on known episodes was intended to prevent participants from building their analogies upon fabricated situations. In turn, restricting the data analysis to whether or not the superficially similar/dissimilar base analog provided by each participant during the first session was in fact retrieved during the generation task allowed the experimenters to assess the effect of superficial similarity on naturalistic analogical

retrieval in a way that remained unaffected by the potential prevalence of either type of base analogs in LTM.

3.1. Method

3.1.1. Participants and design

An initial cohort of 217 students of psychology at Universidad Nacional del Comahue volunteered to participate in the experiment. The final sample consisted of 160 participants who successfully reported an instance of at least one of the four types of situations that were presented to them during the first phase. The degree of surface similarity between the target analogs and their base analogs (two levels: high vs. low) received between-subjects manipulation. The dependent variable was the retrieval of that same analog during the analogy generation task (second phase).

3.1.2. Materials

Four sets of materials were built, each one comprising a target to be used during the second phase and two cues intended to elicit the recollection of natural base analogs during the first phase. The target analogs were simple situations in which a character is about to carry out an action that is somewhat inconvenient, and participants were tasked with generating analogies to known situations in order to dissuade the character from performing her intended action. For example, in one of our sets of materials the main character is said to have discovered passion fruit sorbet, soon attempting to make passion fruit toppings, cheesecakes, daiquiris, etc. Participants had to use analogies to dissuade him from doing so, on the grounds that by eating so much passion fruit he might get fed up with it. For each of the four targets we derived two cues intended to detect a potential base analog during a previous experimental session. The cues for eliciting superficially similar base analogs were derived by means of replacing base objects and relations by their least abstract superordinates (e.g., for the target of Set 1, the cue asked participants for situations in which they had *consumed* some *new food* in excess, with the consequence that they *got disgusted* of it). Derivation of superficially dissimilar cues comprised two steps: (a) the replacement of target concepts by semantically distant concepts at the same level of abstraction, and (b) the replacement of these new concepts by their least abstract superordinates (e.g., for the target of Set 1, the superficially dissimilar cue asked for situations in which participants had *played* a *new game* in excess, with the consequence that they *got bored* of it; Table 3 displays the four targets, with their corresponding cues). For Phase 1, two alternative booklets of cues were built: version “A,” compiling the semantically similar cues of the four sets, and version “B,” grouping the semantically dissimilar cues. Each participant received either version “A” or “B.”

3.1.3. Procedure

The first phase of the experiment was presented as a classroom activity related to the topic of memory, and it was administered during an introductory course of Cognitive Psychology. Whereas half of the participants received the generic descriptions compiled in

Table 3
Materials used in Experiment 2

Descriptions Used for Eliciting a Superficially Dissimilar Base Analog (Phase 1)	Descriptions Used for Eliciting a Superficially Similar Base Analog (Phase 1)	Target Situation and Analogy Generation Task (Phase 2)
<p>1. Have you ever played some new game to excess, with the consequence that you got bored of it?</p>	<p>1. Have you ever consumed some new food to excess, with the consequence that you got disgusted with it?</p>	<p>1. Suppose that a friend of yours, who has discovered passion fruit sorbet, is now interested in preparing his own passion fruit cheesecakes, toppings, daiquiris, etc. Please use analogies to known situations to convince him that by eating so much passion fruit he might get fed up with it</p>
<p>2. Have you ever interrupted the prescribed treatment of an injury, with the consequence that later recovery was much more difficult than it would have been originally?</p>	<p>2. Have you ever stopped making the monthly payments of a debt, with the consequence that clearing the debt later was much more costly than it would have been originally?</p>	<p>2. A friend of yours, who has acquired a costly van with a loan, is tempted to postpone the monthly payments of the plan until he gets a raise in his salary. Please use analogies to known situations to dissuade him from postponing the monthly payments of the loan, because otherwise the debt will get much harder to clear</p>
<p>3. Have you ever learned intuitively how to do something in a rather clumsy way, with the consequence that you ended up never learning the proper way of doing it?</p>	<p>3. Have you ever made a quick fix to an artifact, which was partially successful, with the consequence that you ended up never seeking a fully adequate solution?</p>	<p>3. A friend of yours has discovered that the WC is leaking. He is about to seal it temporarily with <i>suprabond</i>, even though the result will not be completely satisfactory. Please use analogies to known situations to dissuade him from fixing it by himself, since it can prevent him from ever seeking a completely satisfactory solution in the future</p>
<p>4. Have you ever started doing a physical activity despite believing that you would not enjoy it, but with the result that you ended up enjoying it very much?</p>	<p>4. Have you ever attended a meeting despite believing that you would not have fun, but with the result that you ended up having a great time?</p>	<p>4. A friend of yours, who has been invited to a costume party, fears that she might have a hard time there, since she has not been to any costume parties before. Please use analogies to known situations to dissuade her from declining the invitation, since she might find the party more fun than she expects</p>

booklet “A,” the other half received the descriptions included in booklet “B.” Below each of these generic descriptions they were asked to describe own experiences conforming to those descriptions. In case they knew an instance of such a situation, further instructions required them to write down its particulars in detail. Participants were allotted a fixed time of 8 min for each of the four pages of the booklets. Two independent judges analyzed each participant’s answer to determine in which cases the situation reported by a participant was in fact an instance of the received generic situation. Instances in which judges disagreed (14%) were dropped from further analysis. Judges analyzed participants’ responses to version A of the booklets (i.e., superficially similar base analogs) until obtaining 20 participants reporting a base analog for Set 1, 20 participants reporting a source for Set 2, 20 participants reporting a source BA for Set 3, and 20 participants reporting a source for Set 4. Likewise, judges analyzed participants’ responses to version B of the booklets (i.e., superficially dissimilar sources) until 20 participants could be assigned to each of the four sets of materials.

The second phase of the experiment was presented to participants as a study on the use of analogies in argumentation and took place between 10 and 14 days after the first phase. In order to achieve a strong contextual separation between the two phases, it was administered by different experimenters and in a different setting (the Cognition Laboratory) from the previous phase. After receiving a brief instruction on the use of analogy in argumentation—illustrated with an example of a close analogy and an example of a distant analogy—each participant was presented with one of the four possible target analogs. After reading the target situation, participants were asked to provide analogies that could be used to dissuade the protagonist from carrying out his intended action. Participants were allotted 10 min to complete the task and were enforced to base their analogies on real episodes. Once the allotted time had elapsed, participants were given five more minutes to write down all other situations that had come to mind during the analogy generation task, but that were not included among the final analogical proposals. Table 4 summarizes the main phases of the procedure, as well as their methodological purposes.

3.1.4. Data analysis

Among participants’ responses to the generic descriptions received during Phase 1, only one was analyzed for retrieval during the second phase. The chosen base situation was scored as “retrieved” if it appeared either during the analogy generation task itself, or among the responses to the latter question about other situations that were remembered during the analogy generation task. For each participant, two judges received his or her responses to the target situation matched with his or her response to the corresponding cue during Phase 1 (i.e., if a participant in the superficially similar condition received the target of Set 3 during Phase 2, judges received his or her responses to such target, matched with his or her answer to the superficially similar cue of Set 3). In each case, they were to decide whether the situation reported during Phase 1 was included among the analogies reported during Phase 2. Judges agreed in 93% of the cases, resolving cases of disagreement by open discussion.

Table 4

Schema of the different phases of the procedure of Experiment 2, with their methodological purpose

Phases of the Experimental Procedure	Methodological Purpose
1. Participants received descriptions of generic events and were asked to describe autobiographical episodes conforming to such descriptions. These descriptions pointed to either superficially similar or superficially dissimilar analogs of the situation to be presented in the following phase. (10–14 days delay + strong contextual separation)	Detect the availability of an autobiographic episode that was either a superficially similar or a superficially dissimilar analog of the target situation to be presented during the following phase
2. Participants received a definition of analogy and an explanation of its utility for persuasion, illustrated with examples of close and distant analogies	Make participants aware that analogies between situations can be useful for persuasion
3. After reading a situation admitting two lines of action, participants were asked to support one of such actions by analogizing to known episodes. Depending on the condition, the situation was either a superficially similar or a superficially dissimilar analog of the autobiographic episode reported during the prior session	Involve participants in the ecologically valid task of analogical persuasion, and document the use (or not) of the specific autobiographical situation that was detected during Phase 1, and which was analogous to the target situation
4. Participants were asked to write down any other situation that had come to mind during the analogy generation task, describing it exactly as it had been recalled during such task	Identify those cases where the critical episode was in fact retrieved during the prior task, despite not being included among the analogical proposals

3.2. Results and discussion

Superficially similar base analogs were retrieved in 45% of the cases during the analogy generation task (94% of the retrieved cases were included among analogical proposals). In contrast, superficially dissimilar sources were retrieved in 16.25% of the cases (92% of the retrieved sources were included among analogical proposals). In line with results of Experiment 1, results of Experiment 2 showed that superficially similar sources were more easily retrieved than superficially dissimilar ones, $\chi^2(1, N = 160) = 15.562$, $p < 0.001$. This effect held for three of our four sets of materials: 65% versus 20%, $\chi^2(1, N = 40) = 8.286$, $p < 0.01$ (Set 1); 55% versus 20%, $\chi^2(1, N = 40) = 5.227$, $p < 0.05$ (Set 2); and 50% versus 15%, $\chi^2(1, N = 40) = 5.584$, $p < 0.05$ (Set 4). The remaining set yielded a floor effect: 10% versus 10%, $\chi^2(1, N = 40) = 0$, $p = 1$.

These results extend the findings of Experiment 1 to autobiographic episodes, thus failing to support Hofstadter and Sander's (2013) claim that when sources are familiar and meaningful, their retrieval is less constrained by superficial similarity than is typically the case with artificial materials provided by the experimenters. Note that the overall retrieval probabilities obtained in Experiment 2 were lower than those of Experiment 1. We conjecture that this drop in overall retrievability might be due to the fact that the target analogs used in Experiment 1 were larger and more complex than those of Experiment 2. An informal comparison between both experiments in terms of the amount of analogical

proposals that were not based on the critical base analogs revealed a greater number of such proposals in Experiment 2. Based on the competitive nature of memory retrieval postulated by various models (e.g., ARCS, Thagard et al., 1990; LISA, Hummel & Holyoak, 1997), this increased retrieval of non-critical base analogs in Experiment 2 could have inhibited the retrieval of the critical sources to some extent.

4. General discussion

A wealth of experimental studies (e.g., Gentner et al., 1993; Keane, 1987) has demonstrated that superficial similarity plays a central role in analogical retrieval. In recent times, this conclusion has been challenged by a series of naturalistic observations of experts, as they work within their areas of expertise (e.g., Blanchette & Dunbar, 2001; Richland et al., 2004; Christensen & Schunn, 2007). These observations revealed that under specific pragmatics, such as when the reasoners want to communicate ideas to others, they usually select sources bearing little or no superficial similarity with the target situation. In a very influential study, Blanchette and Dunbar (2000) were able to replicate this majority of distant reminders with novice participants working under more controlled laboratory conditions. Blanchette and Dunbar (2000) presented their results as evidence that when participants are allowed to generate their own analogies for realistic situations and tasks, analogical retrieval is less constrained by superficial similarity than was previously thought. These results called into question the ecological validity of more than two decades of experimental research on analogical retrieval, as well as the accuracy of several computational models engineered to simulate and also explain this pattern of behavioral results.

We have pointed out that despite its gain in ecological validity, a series of methodological shortcomings of the production paradigm, as implemented by Blanchette and Dunbar (2000), should preclude interpreting the profusion of superficially dissimilar analogies among participants' proposals as evidence for retrieval processes not severely constrained by superficial similarities. In the first place, the production paradigm did not distinguish between cases of analogical retrieval and cases of analogy fabrication. Second, the pattern of base analogs retrieved by participants may not necessarily reflect the retrieval tendencies of the system, since calculating the retrieval probabilities of superficially similar (or dissimilar) base analogs should also take into consideration the number of available sources that failed to be retrieved, so as to allow taking the quotient between the retrieved cases and all potentially available cases in LTM.

In order to resolve the controversy between the experimental and the naturalistic traditions in the study of analogical retrieval, we designed a hybrid procedure that preserved the ecological validity of the production paradigm implemented by Blanchette and Dunbar (2000), while still retaining the inferential power of the reception paradigm to assess the causal role of superficial similarities in analogical retrieval. In our two experiments, participants read hypothetical situations in which the protagonist was about to carry out an action that could potentially engender negative consequences. Just as in the production

paradigm, their task was to generate analogies that could be used to discourage such line of action—an activity that allegedly promotes a structural encoding of the targets. Also as in the production paradigm, participants were allowed to draw on their own extra-experimental memories, which according to Dunbar (2001), are encoded in a more abstract way than the base analogs typically used during traditional two-phase experiments. Having retained these distinctive advantages of the production paradigm as originally implemented, our studies also preserved those features of the reception paradigm that afford controlling for the potentially distorting effects of analogy fabrication and uneven availability of superficially similar versus superficially dissimilar base analogs in LTM.

In Experiment 1, we controlled for these potential confounds by assessing the retrieval of culturally shared stories (episodes from popular movies) during the generation of persuasive analogies for target situations maintaining high versus low superficial similarity with such episodes. Restricting the data analysis to the retrieval of such episodes afforded calculating the accessibility of these episodes as quotients between the number of retrieved sources and the total number of sources available for retrieval, thus controlling the potentially uneven availability of close versus distant sources. By the same token, the improbability of participants inventing a source analog identical to such movie episodes prevented the experimenters from taking invented sources as cases of analogical retrieval. Finally, by asking participants whether the critical movie had come to mind during the task of generating analogies, we were able to identify those cases in which the critical source analog was retrieved but not included among final proposals.

In line with traditional studies using experimentally provided stimuli, results showed that the probability of retrieving the critical movies after superficially similar cues (70%) was much higher than after superficially dissimilar targets (15%), thus demonstrating a robust effect of surface similarity on naturalistic analogical retrieval. As these results were obtained allowing participants to retrieve their own sources in the service of a target task of alleged ecological validity, they fail to support the increasingly popular thesis that prior failures to elicit distant retrievals constitute an artifact of experimentally implanting the source analogs to be retrieved during tasks of dubious ecological validity (Blanchette & Dunbar, 2000; O’keefe & Costello, 2008).

Even though the procedure followed in Experiment 1 retained all the features identified by the above authors as underpinning the production paradigm’s success in eliciting long-distance retrieval, the fictional episodes chosen as the critical sources might have lacked, by their very nature, the deep personal significance that, for example, Hofstadter and Sander (2013) consider to be at the heart of successful analogical retrieval. According to these authors, the humble analogies that we all make on a daily basis are cut from a radically different cloth from the analogies that are generally studied in the laboratories, in the sense that the former ones involve familiar situations that have been confirmed and reconfirmed over time.

In response to Hofstadter and Sander’s (2013) claim about the key role of familiarity and personal significance in analogical retrieval, in Experiment 2 we set forth to assess the effect of superficial similarity on the retrieval of autobiographical memories. During

the first phase of the experiment, participants were given schematic descriptions of four different situations with the instruction to recall known episodes conforming to any of those descriptions. Less than 2 weeks after Phase 1, participants having reported an exemplar of at least one of the given descriptions were asked to participate in a study on analogical argumentation in a different place and by different experimenters from those administering the previous phase. After reading a realistic situation that, unbeknownst to them, was either superficially similar or superficially dissimilar to the source they had reported during the previous phase, they were asked to draw analogies to real episodes in order to dissuade the main character of that situation from carrying out his intended action. In line with results of Experiment 1, results from Experiment 2 demonstrated a strong and uniform effect of superficial similarities on the retrieval of the autobiographic episodes that had been revealed during the first phase of the experiment. This last result fails to support Hofstadter and Sander's (2013) thesis about the special status of familiar and mundane episodes as candidates for superficially unconstrained analogical retrieval.

Having remedied the insufficiencies detected in the production paradigm as originally implemented, the results of both experiments converge in demonstrating that superficial similarities play a crucial role in the retrieval of naturally encoded sources during analogy generation. Since the personal nature of the sources and the ecological validity of the analogy generation task are precisely those aspects of the production paradigm allegedly underpinning its power to elicit structural retrievals, our results run counter to the claim that the dominance of superficial similarities in retrieval is rooted in the artificiality of the tasks and materials used in traditional experiments.

4.1. Following surface similarities: A feature or a bug?

The results of classic studies like Gick and Holyoak (1980) or Keane (1987) typically elicit a mixture of astonishment and concern (see, e.g., Day & Goldstone, 2012; Loewenstein, 2010). As eloquently conveyed by Gentner et al. (1993, p. 567): "How can the human mind, at times so elegant and rigorous, be limited to this primitive retrieval mechanism?" In terms of evolution, memory retrieval is arguably an older process than analogical mapping, and thus incapable of computing a full-fledged structural match between the target and every representation in LTM.

Regardless of whether memory is capable of computing all structural matches, we will argue that the negative implications of basing retrieval upon surface resemblances have often been overstated, partly as a consequence of regarding surface similarities as being inherently inconsequential. In a classic study, Keane (1987) had college students solve the "radiation problem" (Duncker, 1945), in which a doctor had to figure out how to use rays to destroy a stomach tumor without harming the surrounding tissues. During a prior session, half of the participants had read an analogous story about a surgeon who destroyed a cancer by dividing a strong ray up into weaker beams, and having them converge on the cancer from different locations. The other half of participants had read a story about a general who captured a fortress by dividing his army into small groups, and having them converge on the fortress via different roads, which were mined to explode

when large groups passed over them. The fact that participants retrieved the surgeon story much more frequently than the military story has often been taken to illustrate how memory relies critically on causally irrelevant similarities (e.g., the common presence of tumors, rays, physicians, etc.). But in the world we live in, where most things that look alike are alike relationally as well (the *kind world hypothesis*, Gentner & Medina, 1998), the types of surface similarities that people follow are rarely, if ever, causally irrelevant. Albeit inadvertently, when reasoners include target concepts like *destroy*, *rays*, or *tumor* in the memory probe that will be used for retrieval, they increase the probabilities of being reminded of situations involving concepts that are similar to those of the target, as determined by non-trivial semantic criteria such as whether they belong to the same categories of objects and relations. By so doing, the favored source analogs are likely to be structurally similar to the target, not only at the level of explicitly represented systems of relations (e.g., two cases in which exerting a force against a central target produces collateral damage) but also at a deep, not explicitly represented, physical level (e.g., the fact that in both situations, converging rays have an additive effect at their intersection).³ The tasks we used in Experiment 2, albeit different from problem solving, allow a similar analysis. As an example, consider the task of persuading somebody else that eating too much passion fruit will make her stop liking it. Even though the analogies with (a) somebody who got fed up of ingesting too many lychees, and (b) somebody who got tired of playing too many computer games, might probably obtain similar scores on traditional measures of analogical soundness (e.g., Gentner et al., 1993), it is clear that episodes involving semantically similar entities and relations will bear more predictive power than those involving less similar ones. With the exception of communicative situations—in which the base analog *needs* to be known by the recipient of an analogy—it is hard to see why a bias toward retrieving semantically related situations should be considered a limitation.

With regard to novice-expert differences in analogical retrieval, it has often been claimed that experts are less dependent on superficial similarities than are novices (Faries & Reiser, 1988; Novick, 1988). Gentner et al. (1993) argue that this expert advantage lies not at the level of search mechanisms, but instead at the level of representations, with experts encoding the analogs with more abstract labels which allow retrieval even when surface similarities are lacking. However, even though experts might, in principle, be less dependent on surface similarities, they mostly confine search within their own domains of expertise. As an example, consider you have been diagnosed with a heart disease for which conventional treatments have not yet been tested. Upon a visit to a cardiologist, what would be your reaction if you saw that the bookshelves were packed, not with the usual handbooks of cardiotherapy, but with treatises on military strategy? Just as the chess master has access to about 50,000 stored positions, any expert has acquired her expertise in part by working through many intradomain examples that can now serve as a rich source of analogies that permit efficient problem solving (Feltovich, Spiro, & Coulson, 1997).

To summarize, the types of superficial features used as retrieval cues both by experts and novices most of the time reflect sophisticated semantic criteria such as whether two

concepts belong to specific categories of objects and relations, which in turn increase the probabilities that objects and relations in the retrieved sources will satisfy the constraints of the target domain, even when those constraints are not fully acknowledged by the analogizer. Just as in classic theories of problem solving, which recommend resorting to general heuristics only when more specific procedures are not at hand, we suggest that the proficient analogizer begins by including surface information about the target in the WM probe that will be used for retrieval, and opts for removing target-specific information only if thematically related sources do not come to mind. In recent work (Trench, Olguín, Margni, & Minervino, 2013), we have explored the extent to which the inclusion of surface features as retrieval cues can be consciously relaxed. It turns out that when participants are asked to deliberately search for distant sources, those sources are in fact more likely to be retrieved.

There is, however, a specific kind of situation in which following surface similarities will probably lead us astray: those instructional situations where the learner has to apply general laws or principles (e.g., the concept of *competitive specialization*, Goldstone & Son, 2005) or formal methods for solving algebraic problems (e.g., the probability problems studied by Ross, 1989). We believe that the discussion about the adaptive value of our retrieval tendencies in cases governed by general principles and procedures should be pursued separately from the discussion about whether they are adaptive for dealing with target situations for which thematically closer sources tend to carry more inferential power.

4.2. *Distinguishing analogy generation from analogical retrieval*

The present results demonstrate that a superficial bias governs analogical retrieval not only in laboratory conditions but also during the retrieval of participants' own sources in the service of tasks of indubitable ecological validity. How, then, to reconcile this finding with the results of most naturalistic studies showing a frequent use of distant analogies?

A wealth of anecdotal reports and experimental studies in the literature on creative cognition suggest that the frequent use of distant analogies can be aided by several other mechanisms that do not conform to the standard definition of analogical retrieval, that is, the activation of a base analog from LTM during the processing of a specific target analog in WM. As posited repeatedly throughout the paper, a trivial alternative to scanning LTM for potential analogs consists in inventing possible variations of the target situation. Beyond the obvious usefulness of this heuristic for making a target topic more understandable to a potential listener, there is some evidence that by drawing analogies to hypothetical variations of a target situation, the analogizer can also enhance her own understanding of the target (e.g., Clement, 1988). Another alternative to retrieving one's own sources for a given target is the action of elaborating on sources associated with culturally shared conceptual metaphors (Lakoff, 2008; Lakoff & Johnson, 1980, 1999). As demonstrated by Minervino, López Pell, Oberholzer, and Trench (2009), people find it easy to generate semantically distant metaphorical expressions by means of extending or

reinstantiating conceptual metaphors associated with the target topic, whenever they are available in the culture.

When attempts to generate analogies by means of purely internal processes do not yield appropriate results, suitable analogs are often provided by the environment. During *creative worrying* (Browne & Cruse, 1988; Olton, 1979), even though the reasoner is not being deliberately devoted to solving a given problem, she occasionally interleaves daily activities with lapses of time during which she revisits the problem. Quite frequently, during the lapses in which the target is active in WM, the social or physical environment serendipitously presents the reasoner with relevant analogs, leading to the establishment of a useful analogy.⁴ In pressing situations, such as when analogy researchers are striving to generate experimental materials, reliance on the fertility of the environment can lead to what we would call *analogical foraging*, that is, the process of deliberately screening the external milieu in search of potentially relevant analogs.

We believe that the potential contribution of these alternative mechanisms to the patterns of long-distance analogy generation revealed by naturalistic studies deserves to be studied more systematically, both within natural settings and under more controlled laboratory conditions.

4.3. *The in vivo/in vitro approach to analogical retrieval (and beyond)*

We tend to sympathize with Dunbar's (1999) claim that by observing natural settings, new issues, topics, and ideas can be identified which have not been the focus of the traditional experimental approach, and which have the potential to radically shift the types of questions and theories that a field holds. However, a precipitated move from in vivo to in vitro research may engender serious pitfalls, such as the proclivity to believe that the mere act of reproducing any naturally occurring phenomenon within laboratory settings suffices to endow such phenomenon with a more solid empirical basis. When carried out straightforwardly, in vitro replications can easily inherit the confoundings that characterize their original settings. For example, although the production paradigm represents a clear advantage over naturalistic observations in controlling factors such as the amount of time, interpersonal interaction, or external sources of information available to the reasoners, it shares with naturalistic observations the fact that other variables like analogy invention and uneven availability of sources remain uncontrolled.

Even if one were determined to apply the experimental rigor in its full potential, yet another pitfall of the in vivo/in vitro approach might reside in submitting to experimental manipulation the wrong intuitions about the real factors underlying any naturalistic observation. If the hypothesized causal relationships happened to be reproduced by means of artificial manipulation, one would be more than tempted to assume that the causal factors under consideration *explain* the target phenomenon allegedly revealed by naturalistic observation. But obviously, this need not be the case.⁵

When applicable, a safer way of getting the best of experimental and naturalistic approaches might consist in starting off with a very controlled experimental setup, and progressively incorporating more ecologically valid activities without sacrificing

methodological control. We believe that the hybrid paradigm employed in the present experiments fits well within such an approach.

Acknowledgments

This work was supported by grants PICT 2352 and PICT 2650 from the National Agency for Scientific and Technical Research of Argentina (ANPCyT), grant C108 from National University of Comahue, and grant PIB013 from Open Interamerican University. We thank Arthur Markman, Mark Keane, and two anonymous reviewers for their helpful comments on an earlier draft of the manuscript.

Notes

1. For instance, whereas participants in the production paradigm are roughly equaled in terms of the amount of time, interpersonal interaction, or external sources of information that were available to them, little is known about how these same variables might have gravitated in the generation of the analogies that appeared in newspapers.
2. In a study on problem solving, Chen, Mo, and Honomichl (2004) used a similar procedure to assess the retrieval of popular tales consisting of a problem and its solution. Target problems shared with the base story the goal of weighing a heavy object that exceeds the capacity of available scales. The authors factorially manipulated the similarity between the source story and the target problems in terms of (a) the object to be weighed and (b) the available weighing tools. In line with classic experimental studies, their manipulation of surface similarity affected retrieval. However, the retrieval of the tales in response to targets involving dissimilar tools and to-be-weighed objects was still very high. Regarding our interest in the role of surface similarity in naturalistic retrieval, it is hard to determine whether the high retrieval rate obtained by Chen et al. was a consequence of participants being able to use their own sources, or a consequence of the fact that all targets shared identical goals (weigh a heavy object) and restrictions (scales of insufficient capacity) with the source.
3. The intuition that close analogs convey more relevant information about the target than distant analogs resembles a widely held assumption of the Category-Based Induction literature, which states that if a given category has certain property “x,” then categories with which it shares a higher number of known features are relatively more likely to have property “x” than categories with which it shares a lesser amount of known features.
4. An apparently similar way of stumbling across environmental cues is *opportunistic assimilation* (Christensen & Schunn, 2005; Seifert, Meyer, Davidson, Patalano, & Yaniv, 1995), a process by which the reasoner is reminded of an unsolved problem

during a later encounter with an environmental cue providing its solution. We have not included this process in our review of collateral retrieval mechanisms because we believe that despite being triggered by an external event, the fact that the reasoner retrieves the previously unsolved target problem from LTM justifies the inclusion of opportunistic assimilation as conforming to a stringent definition of analogical retrieval.

5. As an example, in an attempt to scrutinize their naturalistic findings under the rigor of controlled experimentation, Dunbar and Chung (reported in Dunbar, 2001) tested their hypothesis that when both the base and the target are encoded in a manner that stresses their common abstract features, the retrieval of the sources in response to the targets does not necessitate superficial similarities. By means of a traditional reception paradigm, they manipulated how structurally the base and the target were encoded, with superficial similarity kept constant at a very low level. After obtaining that a structural encoding of both the base and the target elicits distant retrievals reliably, the authors interpreted this outcome as a strong indication that what really explains the observed advantage of natural settings is the fact that they promote a more abstract encoding of the sources and the targets than is typical of laboratory tasks and materials. But according to the results of our modified versions of the production paradigm, naturalistic encodings are in no way advantageous in fostering long-distance retrieval, what strongly suggests that the profusion of distant matches obtained in naturalistic settings originates in mechanisms different from analogical retrieval per se. Thus, even though Dunbar and Chung's results might be right in demonstrating that structural encodings can sometimes afford the retrieval of superficially dissimilar sources, by no means do they demonstrate that a structural encoding *is what explains* the observed patterns of analogy generation.

References

- Blanchette, I., & Dunbar, K. (2000). How analogies are generated: The roles of structural and superficial similarity. *Memory & Cognition*, 28, 108–124.
- Blanchette, I., & Dunbar, K. (2001). Analogy use in naturalistic settings: The influence of audience, emotion, and goals. *Memory & Cognition*, 29, 730–735.
- Browne, B. A., & Cruse, D. F. (1988). The incubation effect: Illusion or illumination? *Human Performance*, 1, 177–185.
- Catrambone, R. (2002). The effects of surface and structural feature matches on the access of story analogs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 318–334.
- Chen, Z., Mo, L., & Honomichl, R. (2004). Having the memory of an elephant: Long-term retrieval and use of analogues in problem solving. *Journal of Experimental Psychology: General*, 133, 415–433.
- Christensen, B. T., & Schunn, C. D. (2005). Spontaneous access and analogical incubation effects. *Creativity Research Journal*, 17, 207–220.
- Christensen, B. T., & Schunn, C. D. (2007). The relationship of analogical distance to analogical function and pre-inventive structure: The case of engineering design. *Memory & Cognition*, 35, 29–38.
- Clement, J. (1988). Observed methods for generating analogies in scientific problem solving. *Cognitive Science*, 12, 563–586.

- Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist*, *47*, 153–176.
- Dunbar, K. (1997). How scientists think: Online creativity and conceptual change in science. In T. B. Ward, S. M. Smith, & S. Vaid (Eds.), *Creative thought: An investigation on conceptual structures and processes* (pp. 461–493). Washington, DC: APA Press.
- Dunbar, K. (1999). The scientist in vivo: How scientists think and reason in the laboratory. In L. Magnani, N. Nersessian, & P. Thagard (Eds.), *Model-based reasoning in scientific discovery* (pp. 89–98). New York: Plenum.
- Dunbar, K. (2001). The analogical paradox: Why analogy is so easy in naturalistic settings, yet so difficult in the psychology laboratory? In D. Gentner, K. J. Holyoak, & B. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 313–334). Cambridge, MA: The MIT Press.
- Duncker, K. (1945). On problem solving. *Psychological Monographs*, *58* (5, Whole No. 270).
- Faries, J. M., & Reiser, B. J. (1988). Access and use of previous solutions in a problem solving situation. *Proceedings of the 10th Annual Conference of the Cognitive Science Society* (pp. 433–439). Hillsdale, NJ: Erlbaum.
- Feltovich, P. J., Spiro, R. R., & Coulson, R. L. (1997). Issues of expert flexibility in contexts characterized by complexity and change. In P. J. Feltovich, K. M. Ford, & R. R. Hoffman (Eds.), *Expertise in context: Human and Machine* (pp. 125–146). Menlo Park, CA: AAAI/MIT Press.
- Forbus, K., Gentner, D., & Law, K. (1994). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19*, 141–204.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170.
- Gentner, D. (1989). The mechanisms of analogical transfer. In S. Vosniadou, & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–242). Cambridge, UK: Cambridge University Press.
- Gentner, D., & Forbus, K. D. (1991). MAC/FAC: A model of similarity-based access and mapping. In K. Hammond & D. Gentner (Eds.), *Proceedings of the 13th Annual Conference of the Cognitive Science Society* (pp. 504–509). Hillsdale, NJ: Erlbaum.
- Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*, 45–56.
- Gentner, D., & Markman, A. B. (2006). Defining structural similarity. *The Journal of Cognitive Science*, *6*, 1–20.
- Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*, 263–297.
- 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.
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*, 306–355.
- Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences*, *14*, 69–110.
- Hofstadter, D., & Sander, E. (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. New York: Basic Books.
- Holyoak, K. J., Novick, L. R., & Melz, E. R. (1994). Component processes in analogical transfer: Mapping, pattern completion, and adaptation. In K. J. Holyoak & J. A. Barnden (Eds.), *Advances in connectionist and neural computation theory* (Vol. 2, pp. 113–180). *Analogical connections* Norwood, NJ: Ablex.
- Holyoak, K. J., & Thagard, P. (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: The 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.
- Keane, M. T. (1987). On retrieving analogues when solving problems. *Quarterly Journal of Experimental Psychology*, *39*, 29–41.

- Klahr, D., & Li, J. (2005). Cognitive research and elementary science instruction: From the laboratory, to the classroom, and back. *Journal of Science Education and Technology*, *14*, 217–238.
- Lakoff, G. (2008). The neural theory of metaphor. In R. W. Gibbs Jr (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 17–38). Oxford, UK: Oxford University Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: Chicago University Press.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. Chicago: Chicago University Press.
- Loewenstein, J. (2010). How one's hook is baited matters for catching an analogy. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 53, pp. 149–182). San Diego, CA: Elsevier.
- Markman, A. B., Taylor, E., & Gentner, D. (2007). Auditory presentation leads to better analogical retrieval than written presentation. *Psychonomic Bulletin & Review*, *14*, 1101–1106.
- Minervino, R., López Pell, A., Oberholzer, N., & Trench, M. (2009). A continuist approach to promoting creativity: Generating novel metaphorical expressions through varying conceptual metaphors. In B. Kokinov, D. Gentner, & K. Holyoak (Eds.), *New frontiers in analogy research* (pp. 330–337). Sofia, Bulgaria: NBU Press.
- Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 510–520.
- O'Keefe, D., & Costello, F. (2007). *Analogical retrieval*. Technical report. University College, Dublin, School of Computer Science and Informatics.
- O'Keefe, D., & Costello, F. (2008). A fast computational model of analogical retrieval (and mapping). In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2003–2008). Austin, TX: Cognitive Science Society.
- Olton, R. M. (1979). Experimental studies of incubation: Searching for the elusive. *Journal of Creative Behavior*, *13*, 9–22.
- Richland, L. E., Holyoak, K. J., & Stigler, J. W. (2004). Analogy use in eighth-grade mathematics classrooms. *Cognition and Instruction*, *22*, 37–60.
- Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 456–468.
- Saner, L., & Schunn, C. D. (1999). Analogies out of the blue: When history seems to retell itself. In M. Hahn & S. Stoness (Eds.), *Proceedings of the 21st Annual Conference of the Cognitive Science Society* (pp. 619–624). Mahwah, NJ: Erlbaum.
- Seifert, C. M., Meyer, D. E., Davidson, N., Patalano, A. L., & Yaniv, I. (1995). Demystification of cognitive insight: Opportunistic assimilation and the prepared-mind perspective. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 65–124). Cambridge, MA: The MIT Press.
- Thagard, P., Holyoak, K., Nelson, G., & Gochfeld, D. (1990). Analog retrieval by constraint satisfaction. *Artificial Intelligence*, *46*, 259–310.
- Trench, M., Olguín, V., Margni, A., & Minervino, R. (2013). Automatic and strategic search during analogical retrieval. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1480–1485). Austin, TX: Cognitive Science Society.