COGNITIVE SCIENCE
A Multidisciplinary Journal

Understanding the *What* and *When* of Analogical Reasoning Across Analogy Formats: An Eye-Tracking and Machine Learning Approach

Jean-Pierre Thibaut, Yannick Glady, Robert M. French

University of Bourgogne, Dijon, France

Abstract

Starting with the hypothesis that analogical reasoning consists of a search of semantic space, we used eye-tracking to study the time course of information integration in adults in various formats of analogies. The two main questions we asked were whether adults would follow the same search strategies for different types of analogical problems and levels of complexity and how they would adapt their search to the difficulty of the task. We compared these results to predictions from the literature. Machine learning techniques, in particular support vector machines (SVMs), processed the data to find out which sets of transitions best predicted the output of a trial (error or correct) or the type of analogy (simple or complex). Results revealed common search patterns, but with local adaptations to the specifics of each type of problem, both in terms of looking-time durations and the number and types of saccades. In general, participants organized their search around source-domain relations that they generalized to the target domain. However, somewhat surprisingly, over the course of the entire trial, their search included, not only semantically related distractors, but also unrelated distractors, depending on the difficulty of the trial. An SVM analysis revealed which types of transitions are able to discriminate between analogy tasks. We discuss these results in light of existing models of analogical reasoning.

Keywords: Analogical reasoning; Eye-tracking; Analogy tasks; Search strategies; Machine learning; SVM; Models of analogical reasoning

Correspondence should be sent to Jean-Pierre Thibaut, LEAD-CNRS UMR 5022, University of Bourgogne, Pôle AAFE, Esplanade Erasme, 21000 Dijon, France. E-mail: jean-pierre.thibaut@u-bourgogne.fr

1. Introduction

Analogical reasoning is typically conceived of as a process in which a *base* domain and a *target* domain are compared in order to find relational correspondences between them (e.g., Gentner, Holyoak, & Kokinov, 2001; Holyoak, 2012). Analogies play a central role in many activities and, as such, have been the focus of numerous studies over the years (e.g., Hofstadter & Sander, 2013; Holyoak, 2012; Krawczyk, 2017).

Understanding an analogy is a multifaceted task requiring systematic comparisons between the items in both domains of the analogy problem. Most conceptions of analogical reasoning include the following processes: (1) encoding the items making up the problem; (2) search and retrieval of one relation in memory that connects the two terms, A and B, in the base domain (e.g., "*lives in*" for *bird* and *nest*); (3) mapping of the hypothesized relation holding in the base domain to the target domain, between a C and a D (e.g., "dog" with "doghouse"); and (4) evaluation of the soundness of the mapping (e.g., can both pairs be unified by the "lives in" relation) (e.g., Chen, Honomichl, Kennedy, & Tan, 2016; French, 2002).

At the heart of most analogy models is the mapping process between the base and the target domains. Mapping is the process involved in finding a set of systematic correspondences between the source and the target domain. This means establishing that the relations holding between a subset of objects, events, and characters in the source domain also hold between a subset of objects, events, and characters in a target domain (Holyoak, 2012). Importantly, depending on the model of analogical reasoning, the emphasis will either be on the alignment between entities playing the same role in both domains (i.e., A with C, and B with D) or will involve a generalization of the relation discovered in the base domain and subsequently applied to the target domain. In the latter case, hypotheses must be made regarding which relation(s) in the base domain (i.e., A and B) can be applied to the other domain (i.e., C and D) (for reviews, see Gentner & Forbus, 2011; Gentner et al., 2001; Holyoak, 2012).

The present paper aims to study the temporal dynamics of base-to-target domain mapping by means of eye-tracking, with analogies of various types and varying levels of difficulty. The central issue that we will explore is whether the temporal organization of mapping varies depending on the structure of the analogy task (i.e., scene analogies and proportional analogies), and/or according to characteristics of the domains being compared (e.g., analogy difficulty as measured by the semantic distance between domains or by the type and number of distractors).

1.1. Definition of the search space

The search space of an analogy can vary from well-defined spaces in which the set of potential dimensions is limited, to much more open search spaces in the case of very different source and target domains and/or semantically weakly associated entities in both domains. The matrix completion task (Chen et al., 2016; Sternberg, 1977) is an example of the first. Participants view a matrix of items that share a particular relationship and then select a solution that completes the matrix in a way that is consistent with the relation between the objects in the matrix. Problem complexity is usually defined by a small number of dimensions that

differ in saliency and by the number of transformations that have to be kept active in working memory (see Bethell-Fox, Lohman, & Snow, 1984; Stevenson, Heiser, & Resing, 2013). Its difficulty involves integrating all of the dimensions into a single representation and distinguishing the correct solution from similar alternatives. By contrast, with semantic analogies, the difficulty is a matter of the conceptual distance between the base and target domains and of the association strength between the items involved in both domains. In addition, the targeted relations can be obvious or less obvious, sometimes necessitating semantic rerepresentation when the salience of the relationally best solution is low (e.g., Green, 2016). Finally, the presence of irrelevant dimensions, whether salient or not, also contributes to task difficulty (e.g., Thibaut, French, & Vezneva, 2010a).

1.2. *Computational analysis of the dynamics of analogy-making*

In solving an analogy problem what information should be processed and when? Most research in the field has dealt with interpretations of analogies, their soundness, factors influencing their comprehension, with or without reaction-time data (RTs). Only a few studies have dealt directly with the temporal organization (i.e., the dynamics) of the search through semantic space required to solve an analogy problem (see below). However, computer models have proposed various ways in which the dynamics of solving analogy problems might occur. These computational models more or less explicitly posit a temporal organization of the search for a solution. We will compare four distinct proposals derived from computational models (see French, 2002; Gentner & Forbus, 2011 for reviews)—namely the "*alignment-first*," "*projection-first*," *"parallel terraced scan,"* and "*relational-priming*" models.

In a predicate-argument context, representations consist of predicates and their arguments. In this case, the predicate instantiates a relational structure in the base domain, such as *revolves_around* (earth, sun). In that case, one attempts to align the arguments in the base domain with arguments in the target domain. Gentner and Forbus (2011) call this an "Alignment-first" approach (e.g., SME or ACME). This is derived from the structural alignment hypothesis (Falkenhainer, Forbus, & Gentner, 1989; Markman & Gentner, 1993), according to which the items that compose the base and the target domains are aligned first and inferences are projected from the base pair to the target pair. In the A:B::C:? T(arget) paradigm, this means that one would first align A with C, and would then look for a solution, T (or Ts) that is conceptually aligned with B. This predicts early attention to the A-C pair and to the B-T(arget) pairs.

By contrast, "Projection-first" models (e.g., LISA, Hummel & Holyoak, 1997; DORA, Doumas, Hummel, & Sandhofer, 2008) begin by identifying a set of *relations* that might be relevant to unify the stimuli in the base pair. Once identified, they project this relation from the base pair (i.e., here, the A-B pair) to the target (i.e., the C-T(arget) pair). This predicts early attention to the A-B pair, as participants study the pair for a relation between the two items. This would imply more early AB saccades (i.e., saccades between A and B) than saccades within the target domain or from the base to the target, followed by more attention to C and Target and C-Target saccades. This contrasts distinctly with the alignment-first case predicting more between-domain AC and BT saccades.

The stochastic *parallel-terraced-scan* models (e.g., French, 1995; Mitchell, 1993) constitute the third class of models. These models make no a priori prediction about the order in which items will be aligned, but rather, dynamically discover relations between objects based on the evolving activation levels of the objects and potential relations between them in a semantic network. In this way, the model gradually converges on a coherent structure of the problem upon which an answer is based. Activations in the semantic network change dynamically according to what the program happens to have (stochastically) perceived up to that point. In the dog:doghouse::bird:nest, within and cross-domain relations are discovered correct (*sleeps-in*) or incorrect (*builds*, for bird nest). When nothing matches between the two domains, temperature rises, and the search is extended beyond its normal bounds. Relations with no corresponding relation in the base domain lose activation and the activation of relational match *sleeps-in* and its associated arguments increase with dynamics that can differ from those of the "alignment-first" or "projection-first" approaches.

Importantly, most conceptions of analogical reasoning, in particular the three mentioned above, involve in one way or another the idea of a selection or discovery of relations among a set of possible relations, allowing us to progressively converge on a solution, analogically correct or not. This is one of the central tenets of most models of analogy-making. They converge on the idea of a lessening of activation for unrelated, local, matches in favor of an activation/construction of a relational structure, generally written in a predicate-argument format.

A fourth view, the *relational priming* model, proposed by Leech, Mareschal, and Cooper (2008), gives no central role to mapping in the development of analogies. According to this model, children first study the A-B pair. Then, the relation found between A and B (the authors appeal to the retrieval of a relevant transformation between A and B) (e.g., *cuts* for *knife* and *bread*) *automatically primes* the retrieval of a relationship between C and the target item into which it is transformed. This model predicts that participants study the A-B pair first, they then turn to C and the solution set, much like the projection-first model. However, in contrast to the projection-first model, the relational-priming model's solution to the problem is found through *priming*, which involves no systematic comparisons between pairs of items in the target domain and mostly ignores distractors. The solution has been primed by the original relation found in the base domain and is directly applied to C. In other words, there is no notion of mapping, or even of an active search, in this model.

Finally, most models include an evaluation of the solution at the end of the process, determining to what extent the inferences discovered are relevant to the context at hand. The precise temporal organization of this evaluation process remains an open question. In principle, an evaluation would be associated with "checking" transitions by going between the target and base domains, in other words, checking the compatibility of their solution in the target domain with the relation they found in the base domain. The evaluation would generally also involve comparisons between the target and the semantic distractors. An open empirical question is whether these processes take place during the entire trial or at the end of the trial.

2. Eye-tracking in analogical reasoning

2.1. The temporal dynamics of analogical reasoning

Thibaut, French, and Vezneva (2010b) characterized analogical reasoning as a search in a space that is dynamically constructed as comparisons proceed. By definition, analogical reasoning involves multiple sources of information and various comparisons within and between the domains making up the problem and their integration into a consistent relational structure. This means that perceptual or semantic similarities or local relations that, initially, seemed important might be discarded during the construction of a relational system that best unifies the two domains (see the systematicity constraint, Gentner, 1983). Eye-tracking technology can allow us to identify differences between types of analogies, between levels of difficulty, whether a child or an adult is solving the problem (e.g., Thibaut & French, 2016), and even whether or not a correct answer to the problem will be given. This is largely because looking positions and times are highly correlated with the independently assessed informativeness of regions within a scene (e.g., Rayner, 2012).

The use of eye-tracking techniques that we have developed (French, Glady, & Thibaut, 2017; Thibaut & French, 2016) and that are used to analyze the results from the two experiments reported in this paper will allow us to show that time-course predictions based on the above computational models do not, in general, present a fully accurate picture of the dynamics of how analogy problems are actually solved.

2.1.1. Eye-tracking contributions to analogical reasoning

Previous eye-tracking studies can be compared along a number of dimensions: format of the analogies (semantic analogies or matrices), types of distractors used (semantically or perceptually related), global analyses on entire trials or slices of trials, age of participants (i.e., children or adults), level of problem difficulty, and a comparison of projection-first, alignment-first, and relational-priming views of analogical reasoning. Bethell-Fox et al. (1984) were among the first to investigate participants' strategies based on eye movements. They studied Raven geometrical matrices and manipulated problem difficulty by modifying the number of transformations, the similarity to item C in the matrix, and the number of alternatives. They hypothesized that these factors would elicit alternative strategies. In this respect, their paper was able to distinguish between a constructive-matching strategy, which is analogous to the projection-first strategy (a first analysis of the matrix is followed by an exploration of the solutions), and a response-elimination strategy that consisted of successive back and forth explorations between the matrix and each of the alternatives. They showed that the manipulated factors influenced participants' strategies (see also Mulholland, Pellegrino, & Glaser, 1980, in a *true-false* judgment task). However, they did not address the temporal dynamics of solving the problems.

Using a similar matrix task in a developmental context, Chen et al. (2016) compared the strategies of 5- and 8-year-old children. Differences in performance between age groups resulted from not using optimal processing strategies. Children who employed optimal processing strategies early on were more likely to correctly solve subsequent

problems in later phases. One interesting feature of the study is the authors' equating types of transitions with specific processing strategies, such as item encoding, rule integration, and so on. However, these authors provide only global measures and nothing about a moment-to-moment analysis of the strategies (i.e., when encoding, or rule integration, takes place during solving). With adults, Hayes, Petrov, and Sederberg (2011) applied a novel scanpath analysis in order to capture statistical regularities in eye-movement sequences in the Raven's Advanced Progressive Matrices. They identified two principal components predicting individual scores, the first being the row-by-row scan, and the second toggling toward the response area. The authors interpret the row-by-row strategy as a clue to constructive matching and toggles as a clue to response elimination. The authors did not analyze when these toggles took place but note that this can be done by contrasting the beginning and the end of the trial, an approach we follow. This is important, though, because toggles might also appear in a constructive matching strategy, when participants compare the options with the regularities they found in the base.

Gordon and Moser (2007) conducted an eye-tracking study of analogical reasoning in adults using scenes from Richland, Morrison, and Holyoak (2006). Participants initially focused on the "actor-patient" pair in the source scene (a dog chasing a cat, see Fig. 4a, below, top panel). They then looked for the solution in the target image (a second actor-patient pair: a girl chasing a boy, Fig. 4b, below, lower panel). Significantly, the authors also studied saccades involving the distractors (i.e., saccades between the actor and the perceptually similar distractor in the target pair) and showed that these saccades also occurred after saccades toward relational matches. This suggests that participants did not systematically process object matches before relational matches, which, according to the authors, contradicts Rattermann and Gentner's (1998) claim "that object matches are generally computed before relational matches" (p. 471). However, the study did not provide a systematic analysis of the saccades between the source and the target scenes. It is thus difficult to test the projection-first versus the alignment-first hypotheses. Finally, eye-movement analyses focused on data collected during a 10 s study period *before* the arrow pointing to a source object was introduced in the upper scene. Without the arrow, participants might have engaged in a less targeted search, identifying characters and stimuli, potential relations, but without the attentional weights that might be elicited by the targeted (arrow-based) role.

Thibaut and French (2016) used eye-tracking to study the development of analogical reasoning, from 5 years of age to adulthood (5-, 8-, 13-year-olds and adults) with classical proportional analogies (A:B::C:? paradigm). In order to study the "temporal dynamics" of solving problems, they split each trial into three identical time slices and analyzed the distribution of gazes toward areas of interest (AOIs) and transitions between AOIs. One major result was the significant differences between age groups in the temporal distribution of gaze profiles. Crucially, adults seemed to follow mostly a projection-first strategy, whereas children started by studying the C item and organized their search around it, which can be characterized as an "undirected" search strategy. At the onset of the trials, children had significantly fewer AB transitions than adults and more of other types of transitions. Adults also seemed to be more engaged in solution monitoring, since, by the end of the trial, they had more saccades from the solution set to the AB pair. Children's results were interpreted in terms of difficulty

in inhibiting the main goal of the task, that is, "what goes with C," so that they could focus on the AB pair (see Glady, French, & Thibaut, 2017). Overall, the results showed that adults first analyzed the AB pair and then applied that relation to the target domain. Thereafter, they checked their solution by continuing to look at the distractors and returning to the source domain. In sharp contrast, children tended to focus early on the C item, ignoring, or at least attaching less importance than adults to the AB pair. Both for adults and children, our data revealed little, or no, evidence of AC and/or BT alignments, the key to alignment-first models. This lack of AC and BT saccades has been confirmed by Vendetti, Starr, Johnson, Modavi, and Bunge (2017) in their work on similar proportional analogies (see also Starr, Vendetti, & Bunge, 2018). The latter authors contrasted three strategies, including the projection first and the alignment first. Each was implemented by a different algorithm essentially based on a different subset of early gazes and transitions. The authors followed a "winner-take all approach in which a trial was classified as a particular strategy" (Vendetti et al., 2017, p. 4) depending on its score. In both papers, Starr and colleagues' results favored the projection-first strategy. Hence, by construction, only early gazes were considered, not the temporal dynamics of the entire trial.

In short, eye-tracking data show that adult participants generally favor a mostly projection-first approach. However, with the exception of Gordon and Moser (2007) and Thibaut and French (2016), most studies do not consider the temporal organization of the trial, which we will try to capture with analyses of the beginning, the middle, and the end of a trial (see results, below).

3. Goals and predictions

Our two experiments tackle one main question—namely, how participants organize their search as a function of analogy characteristics. We compare search-strategy adaptations as a function of analogy type (i.e., scene analogies and proportional analogies) and problem difficulty (i.e., a greater number of potential relations between items, weaker semantic relations between items, etc.).

One way to study the effects of the difficulty of semantic analogies in a well-controlled setting is to refer to semantic distance. The semantic distance between domains has been of particular interest in analogical reasoning research. For example, Vendetti et al. (2012) assessed participants' validity judgments of A:B::C:D analogies with near-domain (A and C, and B and D, belonging to the same domain) and far-domain analogies (the source and target domains involving different conceptual domains). They described a general decrease in correct responses in far-domain analogies. Semantic distance has also been shown to affect the activation of brain areas and the dynamics of brain processes during analogical judgment (Green, Kraemer, Fugelsang, Gray, & Dunbar, 2010; Kmiecik, Brisson, & Morrison, 2019).

3.1. Exploring the temporal dynamics of analogy-making with semantic analogies

Former eye-tracking studies with analogies had limitations in the sense that they considered only one type of analogy, most of them matrices based on a set of well-defined dimensions.

By construction, matrices display logical progressions across stimuli and heavily rely on the identification of clear, differentially salient, dimensions. Semantic analogies, by contrast, are defined along semantic spaces that cannot a priori be described with a finite set of dimensions: an infinite number of descriptors can be applied to any situation, surface semantic features, or highly sophisticated interpretations (Hofstadter & Sander, 2013; Murphy & Medin, 1985). Previous studies have not referred to clear moments of a trial with semantic analogies in which *difficulty* was a key variable in the task (French et al., 2017; Gordon & Moser, 2007; Thibaut & French, 2016).

Most eye-tracking data are compatible with a projection-first hypothesis, showing that the base domain is studied first and the results of this exploration are then applied to the target domain. Our paper considers various task formats in different analogical mapping formats, specifically standard proportional analogies and scene analogies. We analyze the time of course of comparisons taking place until a decision is made, and how distractors are processed and rejected as a function of analogy difficulty.

We will manipulate difficulty in several ways, and use techniques we have developed (French et al., 2017) to analyze the predictive and classification power of sets of transitions between AOIs. The idea is to use machine learning classification algorithms to study the predictive dimensions of participants' scanpaths, aiming at identifying "transition profiles" (i.e., subsets of saccade types) that might distinguish complex trials from simple ones, correct trials from errors, or predict the output of a trial (correct or error) in the first third of the trial.

3.2. Task specificity

Do participants adapt their search strategy to the specifics of the analogy tasks? Given the preeminence of the projection-first strategy in existing studies, our main hypothesis was that the projection-first strategy is driven, at least in part, by the particular analogy format, with the layout of proportional analogies implicitly encouraging participants to first seek a unifying relation that would drive the search for the compatible target item. By comparison, scene analogies provide participants with structured scenes with oriented relations in which the identification of equivalent items becomes the central issue. In terms of a predicate-argument structure, (e.g., *chase* (cat, mouse)), the participants' task is to find the first argument of the predicate, *chase*, in the target image that plays the same role as the first argument of *chase* in the source image. Complex (difficult) analogy problems, in comparison with simpler ones, require more evaluation processes, which is often thought of as an alignment of predicates playing the same role.

Our hypotheses are as follows:

Hypothesis 1 is divided into three options—namely:

- Hypothesis 1a, based on prior studies, predicts a general pattern of exploration of the base domain (i.e., A, B) and then the target domain (i.e., C, T). Empirical data should confirm this pattern across analogy formats and complexity levels.

- Hypothesis 1b predicts a higher number of alignment transitions (i.e., AC and BT) with more complex analogies or with the scene format (Experiment 2). Indeed scene might involve more search for equivalence between items (see above).
- Hypothesis 1c predicts a mixture of projection and alignment. Under this hypothesis, participants would analyze in a back-and-forth manner the A-B pair and C-T and C-D pairs. The A-B relation would then be applied to the items in the target domain. Once this is done, they might compare A and C, or B and D as a means of verifying their answer.

Hypothesis 2. Most models describe analogical reasoning in terms of a progressive convergence toward the relational, analogical solution (see French, 2002, Gentner & Forbus, 2011) and assume in one way or another, that local, irrelevant, matches are disregarded by the end of the process (see Gordon & Moser, 2007; Rattermann & Gentner, 1998, for discussions).

- Hypothesis 2a predicts that participants will focus less on distractors at the end of the trial than at the beginning
- Hypothesis 2b predicts that distractors will be compared to the other stimuli less often in easy (or simple) trials than in complex trials because the solution comes more easily to mind and, as a result, unrelated distractors can be ignored more quickly than other stimuli. The general notion of convergence toward a correct solution predicts that by the end of the trial, when a correct solution is chosen, most distractors should be deactivated.

3.2.1. Predicting analogy type, complexity, and performance: A machine learning approach

The second main goal of the paper was to use machine learning classification algorithms to identify a subset of item-to-item transitions (saccades) that might distinguish complex trials from simpler ones and correct trials over erroneous trials (see French et al., 2017 for a description of these techniques). By looking at various "transition profiles" (sets of saccades), we are able to predict with a high degree of accuracy whether the problem being solved is complex or simple. We used a simple and powerful classification technique, a support vector machine (SVM, Vapnik, 1999, combined with a leave-one-out cross-validation, LOOCV, see below). Using this technique, in French et al. (2017), we were able to predict whether an adult or a child was solving a particular analogy problem and, this, in the first third of the trial, based only on the distribution of the transitions (saccades). Using an identical approach in the present paper, we considered how well various transition profiles predicted the difficulty of a problem under scrutiny (see the Results section for more details).

Experiment 1 consisted of proportional analogies with words. We divided the analogies into two classes ("simple" and "complex") according to their difficulty.

Experiment 2 was similar to the first experiment, with the difference being that we used scene analogies with two levels of difficulty ("easy-simple," one relation, and "complex," two relations).

4. Experiment 1: Word analogies

The goal of this experiment was to test whether the level of complexity and the type of distractors (semantically related or unrelated distractors) would influence the time course of analogical reasoning. Previous experiments have manipulated the difficulty of analogy problems by changing the number of distractors, the number of potential solution options, the saliency of feature dimensions (the typical case being matrices of the Raven type, Bethell-Fox et al., 1984), semantic distance (Green et al., 2010), the presence of cross-mappings (Gentner & Toupin, 1986), and so on. For example, with semantic analogies, Green et al. (2010) (see also Kmiecik et al., 2019) studied the neural activation dynamics of brain processes during analogical-judgment tasks, with more errors for semantically distant analogies (see also Bugaiska & Thibaut, 2015) or analogies constructed around weakly or strongly semantically associated pairs (see Vendetti et al., 2012; Vendetti, Wu, & Holyoak, 2014). Green et al. (2010 and 2012; see also Bendetowicz, Urbanski, Aichelburg, Levy, & Volle, 2017; Hobeika, Diard-Detoeuf, Garcin, Levy, & Volle, 2016) have shown that frontopolar activation increases when the semantic distance between the A-B and C-D pairs in a "true/false" verbal analogy problem was increased. In general, an analogical transfer is more difficult when the conceptual domains involved are remote rather than close (Gick & Holyoak, 1980; Keane, 1987).

However, how analogy difficulty correlates with search strategies remains an open question. As summarized above, Thibaut and French (2016) followed a developmental approach. However, their adult participants saw only problems that could be solved by 6-year-olds, which did not allow the authors to directly test the effect of task difficulty. In the present experiment, our general hypothesis predicts that semantically more complex trials would produce different search strategies than simpler problems. Indeed, they may evoke several highly associated words (in the case of word analogies) that are not necessarily relationally consistent solutions and need to be inhibited in order to rerepresent the pair in terms of novel relations (see Collins & Loftus, 1975; Murphy, 2002; Steyvers & Tenenbaum, 2005, for discussions of the notions of semantic networks).

In this experiment, the analogy difficulty was defined in terms of semantic relatedness. Simpler trials were trials in which the unifying semantic relation was relatively obvious compared to Complex trials. For example, a Simple trial such as cow:milk::hen:? has the relation *produces* and the solution, "egg," can be compared to the more complex relation *is-a-type-of* for violence:activity::gloom:? which was rated as complex (see Materials). Several hypotheses can be made regarding the time course of the trials in these cases with differing levels of difficulty.

The general predictions above would be that we should observe the same projection-first profile in both complexity conditions. However, Complex-problem items *might* elicit more alignments than Simple problems because a larger number of alignments might result from the difficulty required in establishing which stimuli are semantically equivalent while playing the same role in the two domains. Thus, a key purpose of the present experiment is to examine how these two strategies could combine.

Second, it has been claimed that object matches (i.e., surface-similarity) are processed before relational matches (Goldstone, 1994a, 1994b; Rattermann & Gentner, 1998). In the

present study, strongly semantically associated distractors will induce surface-similarity matching. Our Hypothesis 2, above, predicts that in Complex trials, participants will produce more transitions involving distractors (e.g., Semantic distractor transitions [CSemDis]) *at the beginning of the trial* than in Simple trials. This is because for Complex trials participants begin by considering semantically related options and because the semantic space is more open. Later on, following most models, we predict that distractors should receive much less attention in both conditions. We also predicted that the early imbalance observed in Simple trials in favor of AB compared to other transitions (more AB transitions than other transitions) should decrease in the complex case. Indeed, less obvious relations should elicit a more systematic search between C and solutions in order to find a solution when one is not *immediately* forthcoming.

Hypothesis 2 also predicts few gazes toward semantically unrelated distractors at any moment in a trial, except at the beginning when participants start to explore the available options. One exploratory question here is whether participants might check less obvious or less plausible solutions, especially in complex trials.

5. Methods

5.1. Participants

Participants were 20 students at the University of Burgundy ($M = 23.8$ years; $SD = 4.2$; from 17 to 35 years). They participated voluntarily, gave their informed consent, and were unaware of the goals of the experiment.

5.2. Materials

The task consisted of 22 trials (two practice trials and 20 test trials) of a verbal A:B::C:D task. The test trials consisted of 10 Complex trials and 10 Simple trials. The presentation of Complex and Simple trials was random. Two practice trials were introduced before the 20 test trials.

Each trial was composed of eight words written in black ink on a white background, corresponding to the A, B, and C terms of the analogical problems, and five potential solutions. The solution set was composed of the Target (T), two semantically related-to-C distractors (SemDis), and two unrelated distractors (UnDis). Each word was presented in a black frame (220×220 pixels). The A, B, and C terms were presented in a row at the top of the screen along with an empty black frame (for the stimulus as the answer), and the five words composing the solution set were displayed in a row at the bottom of the screen (Fig. 1).

Twelve university students assessed trial complexity prior to the start of the experiment. They were asked to solve the different problems and to evaluate the difficulty/complexity of each problem on a 1–7 scale. Complex trials were rated as significantly more difficult ($M = 3.9$; range 3.5–4.6) than simple trials ($M = 1.2$, range 1.1–1.3; two-sample related t-test: $t(22) = 23.2$, $p < .001$, $\eta^2_p = 0.961$).
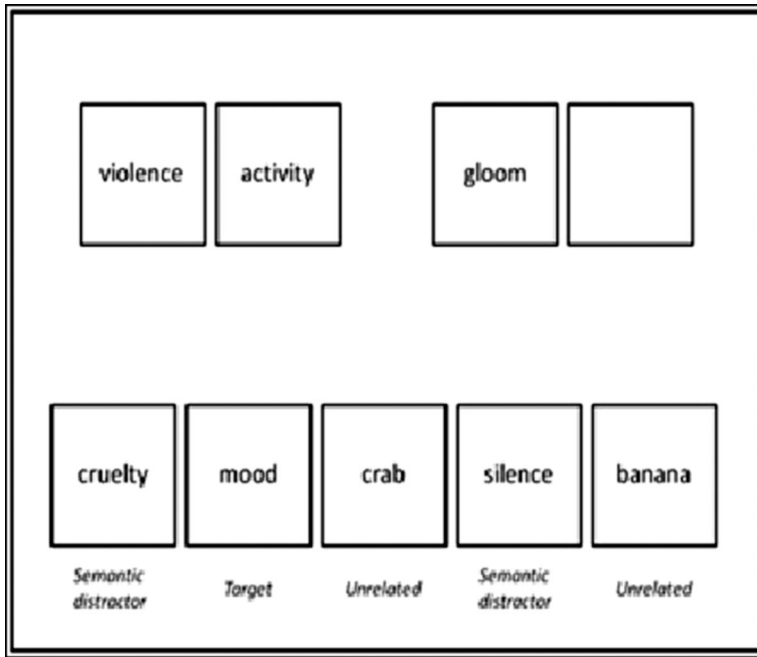
Fig. 1. Example of the display used in Experiment 1.

The task was presented on a Tobii T120 eye-tracker (17″ TFT monitor, resolution: 1024×768) with an E-Prime (version 2.8.0.22) experiment embedded in a Tobii Studio (version 2.1.12) procedure to record participants' eye movements.

5.3. Procedure

Test sessions took place in a dedicated soundproof room. Each participant was tested individually. The distance between each participant's face and the screen was approximately 70 cm. The task started with eye-tracking calibration. The participants were then tested in the analogical reasoning task. The eight words constituting an analogy were displayed simultaneously (Fig. 1) and participants were given the following instructions during the first training trial: "Here are two words [pointing to A and B]. They go together well. Can you see why these two [A and B] go together?" Once the participant had given a relation, the experimenter confirmed it or corrected it and continued: "OK! Do you see this one [pointing to C]? What you have to do is to find among these five words [pointing to the solution set] the one that goes with this one [C] in the same way as this one [B] goes with this one [A]. So, if these two [A and B] go together because [giving the relation between A and B], which one goes with this one [C] in the same way?" When participants had given an answer, the experimenter asked them to justify their answer and gave corrective feedback when necessary. The test trials followed. Participants received no further instructions or feedback. Eye-tracking data

were recorded when the presentation of the problem started and stopped when an answer was given.

6. Results

Before analyzing the time course of saccades (transitions) between objects, we checked whether Complex trials were, in reality, more difficult than Simple trials. The mean rate of correct answers was significantly lower in Complex problems than in Simple problems, $t(19) = 4.9, p < .001, \eta^2_p = 0.558$, with $M = 100\%$ and 79.5% correct for Simple and Complex trials, respectively. All errors were semantic distractor choices. Complex trials were also solved significantly slower than Simple trials, $t(19) = 9.92, p < .001; \eta^2_p = 0.838$, with $M = 4794$ and 12,748 ms for Simple and Complex trials, respectively.

6.1. Eye-movement analysis

We rejected trials in which more than 50% of the gaze time was not recorded. Preliminary analyses we did when we started this project revealed that beyond a given percentage, loss of data did not affect the overall results. For example, using 30% of the data or 80% gave virtually identical results, that is, were perfectly correlated. With this criterion, only two problems were discarded from the entire data set. We focused on two complementary measures of interest, gaze duration and the number of saccades. Gaze duration (or looking times) for AOIs (i.e., gazes toward the items themselves) tells us which items are attended to and for how long they are attended to while solving the problem and witnesses the depth of processing of the item. Saccades between items (or switches, or transitions, i.e., when a participant goes, for example, from item A to item C) tell us which items are compared and can be interpreted as an attempt to find a relation between X and Y (see Duchowski, 2007). These two measures correspond to different aspects of the search, insofar as a participant can study an item for a large amount of time without comparing it with other items. Gaze duration, unlike transitions, tells us nothing about which items are compared during trials.

In order to compare Simple and Complex trials, we first focused on AOIs (gazes) and analyzed their distribution throughout the trial, a distribution which is expected to differ in the two types of trials. Then, we will move to saccades (i.e., "transitions"). In both analyses, we divided all trials into three equal slices (i.e., each slice being 1/3 of the total length of the trial), in order to capture differences in the temporal dynamic of Simple and Complex trials. Indeed, most studies do not analyze the temporal dynamics of trials. In preliminary studies that later led to Thibaut and French (2016) and French et al. (2017), we started with a finer five-slice analysis which gave overly complex results (interactions) essentially similar to the ones reported here. Here, a three-slice approach allows us to separate early explorations of the semantic space from late explorations in the trial which can be interpreted as decisional. Compared with diachronic analyses, such as scanpath analyses, which are supposed to find paths participants might systematically follow (see French et al., 2017; Hayes et al., 2011;

Le Meur & Baccino, 2013, for reviews and discussion), our analyses remain, to some extent, synchronic even though they focus on three different moments.

6.1.1. *Gaze duration analysis (AOIs)*

In this analysis, we focused on the proportion of time spent on each AOI depending on the complexity of the problem being solved. We analyzed AOI gaze duration, and the time course of gazes, splitting trials into three time slices. AOIs were of six types A and B, C and T, SemDis and UnDis (see above). We averaged the two unrelated distractors (UnDis) which played the same role in the design. The same was true for the two semantically related (SemDis) distractors. Then, we regrouped these six AOI types in four stimulus classes. A and B, the base domain, were averaged as one data point (i.e., A&B). The same is true for C and T(arget) (as C&T) which are the analogous stimuli in the target domain. The third class consisted of semantically related distractors (SemDis). The fourth class was comprised of unrelated distractors (UnDis), for which models predict that they should receive little attention, particularly toward the end of the trial. The resulting values were then transformed so that the sum of the four classes in a slice would amount to 33.33% (and so the three time slices would amount to 100%).

A three-way repeated measure ANOVA, with Stimulus class (A&B, C&T, SemDis, and UnDis), Difficulty (Simple and Complex), and Slice (first, middle, and last) as within-subject factors, was performed on the proportions of time spent on the four categories of AOIs to assess the temporal dynamics of fixations. The analysis revealed an expected main effect of Stimulus class, $F(3, 57) = 22.56$, $p < .0001$, $\eta^2_p = 0.54$), an interaction between difficulty and stimulus class, $F(3, 57) = 7.1$, $p < .001$, $\eta^2_p = 0.27$), an interaction between Slice and Stimulus class, $F(6, 114) = 100.37$; $p < .00001$; $\eta^2_p = 0.84$).

The *most important result* was the significant interaction between the three factors, $F(6, 114) = 22.43$, $p < .0001$, $\eta^2_p = 0.54$). Tukey HSD on individual slices revealed the following pattern ($p < .01$). *Slice 1* revealed that the simple condition had a significantly higher proportion of A&B and a lower proportion of fixation times for both types of distractors (SemDis and UnDis) than the complex condition. In *Slice 2*, A&B was longer for the complex than in the simple condition. In *Slice 3*, there was no difference between the simple and the complex condition.

The within-condition analyses showed that in the simple condition, in Slice 1, A&B was significantly longer than C&T as well as both types of distractors, and C&T was longer than both types of distractors. In the complex condition, however, A&B was longer than the others but there was no significant difference between C&T and the two distractor types, revealing a flatter distribution of AOIs in this condition. In Slice 2, in the simple condition, C&T and both types of distractors had longer looking times than A&B, showing that participants were done with A&B. There were more SemDis than C&T, suggesting thorough explorations of semantically related distractors at this stage. In contrast, there was no significant difference between the four types of AOIs in the complex condition (except C&T > SemDis), suggesting a more balanced exploration of the four stimulus types in this condition. In slice 3, the proportion of C&T and SemDis gazes was higher than the proportion of A&B gazes. In addition, C&T was longer than both types of distractors in the simple case. In comparison, in the
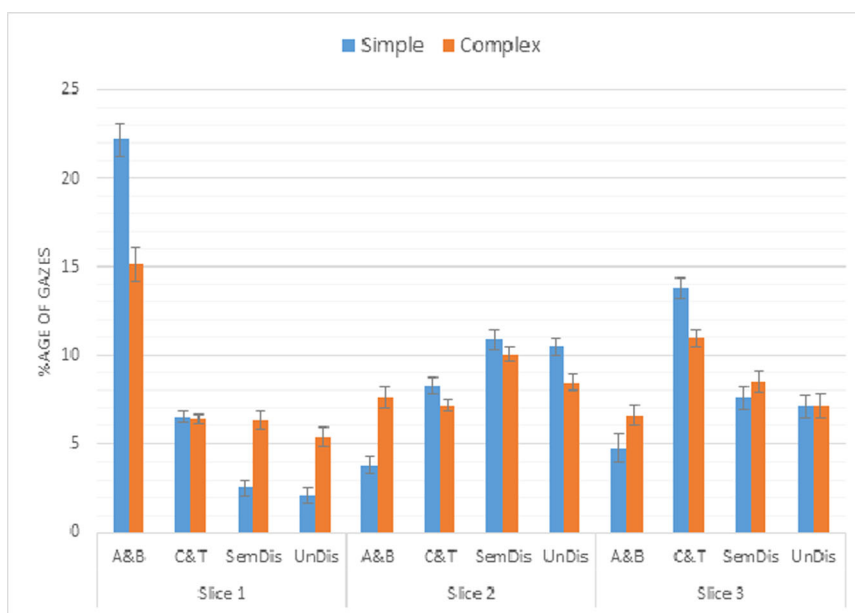
Fig. 2. Mean percentage of fixation time of each Type of Stimulus (AOIs) in the first, middle and last slice in Simple and Complex trials (error bars represent SEM). We collapsed A and B, C and T, respectively, in one score.

complex case, C&T was longer than A&B and UnDis, but not than SemDis, suggesting that in this condition, participants were still struggling with the semantic distractors. This profile will be confirmed by the analysis of the transitions (see Appendix A for the list of confidence intervals).

This is a fascinating pattern of interaction since, as Fig. 2 shows, overall, participants have a "flatter" search pattern in the complex condition in which they look significantly more at both types of distractors at the beginning of the trial than in the simple condition. In the simple condition, the second and third slices show that participants rapidly discarded A&B (i.e., understood the relation between them) and devoted a significantly greater amount of time to C&T and the distractors. The pattern is consistent with the hypothesis that they continue to check alternative solutions (Distractors) until the end of the trial, when they finally converge on the analogical solution. In the complex condition, the *flatter distribution* in the three slices suggests that participants test and check *all* possible solutions against A&B during the entire trial. We speculate that participants explored the entire space of solutions, including irrelevant ones, from the beginning of the trial to its end more thoroughly than in the simple condition. This is consistent with the idea that participants tested other interpretations of the AOIs when the first analysis of AB does not lead to an obvious solution.

6.2. Transitions

We focused on a subset of five sets of transitions which have theoretical meaning (French et al., 2017; Thibaut & French, 2016). This subset was composed of the following transitions

and sets of transitions. When several transitions were included, the resulting "transition" was the average of the component transitions.

Transitions are written in the following format, AB or CT are transitions between A and B or between C and T, and so on. AB are the average of saccades from A to B and B to A. We created the five following transition types.

-{AB}, transitions between A and B and vice versa, within the source domain

-{AC&BT}, an average of AC and BT switches. This represents alignments of equivalent stimuli in both domains that is A with C and B with T,

-{CT}, transitions between C and T, in the target domain

-{CSemDis, TSemDis}(hereafter C&T_SemDis), the average of CSemdis and TSemdis, are transitions between C or T(arget) toward the semantic distractors, SemDis, that have been shown by Thibaut and French (2016) to be important because they indicate how participants reach the solution through comparisons between C and the semantically related stimuli, the semantic distractors, and the target,

-{CUnDis, TUnDis, SemDisUnDis}(hereafter, C&T&SemDis_Undis), which is the average of the transitions between C, Target, and SemDis toward the unrelated distractors, UnDis. This measure aggregates the stimuli which are associated with C and the semantically unrelated distractor, UnDis. It contributes to test a common prediction of most models that Unrelated distractors would be quickly discarded from participants' search space, compared to semantically related stimuli (like C&T_SemDis).

The first three sets of transitions (AB, AC&BT, and CT) are crucial in determining whether participants follow projection-first, constructive strategies (AB then CT(arget)), or alignment-first strategies (a large number of AC and BT) or a combination of both, depending on the moment of the trial. The last two transition types, C&T_SemDis and C&T&SemDis_Undis, will tell us when participants are focusing on T and the semantic or unrelated distractors. The four models presented earlier predict that transitions to the unrelated distractors, that is, Undis, should remain rare, especially at the end of the search, because the system converges on the correct solution.

We ran a three-way repeated-measure ANOVA on the log transformation of the number of transitions, which were not normally distributed, with Transitions (AB, CT, AC&BT, C&T_SemDis, and C&T&SemDis_Undis), time Slice (first, middle, and last), and Condition (Simple and Complex) as within-subject factors. Results revealed a main effect of complexity, $F(1, 19) = 45.01$, $p < .0001$, $\eta^2_p = 0.70$; a main effect of slice, $F(2, 38) = 27.68$, $p < .00001$, $\eta^2_p = 0.59$; a main effect of transitions, $F(4, 76) = 211.28$, $p < .00001$, $\eta^2_p = 0.92$. It also revealed an interaction between complexity and transitions, $F(4, 76) = 12.50$, $p < .0001$, $\eta^2_p = 0.397$; slice and transitions, $F(8, 152) = 66.44$, $p < .0001$, $\eta^2_p = 0.78$. The most important result was the significant interaction between these three factors, Type of Transition, Condition and Slice, $F(8, 152) = 12.89$, $p < .0001$, $\eta^2_p = 0.40$ (Fig. 3).

In the post-hoc comparisons (Tukey HSD), we retained only significant differences below $p < .005$ (see Appendix A for the confidence intervals per condition).

They revealed that, in *Slice1*, the comparison between simple and complex analogies revealed higher rates of CT and C&T&SemDis_Undis transitions in the complex than in the simple problems ($p < .0001$). Within-condition comparisons revealed higher rates of
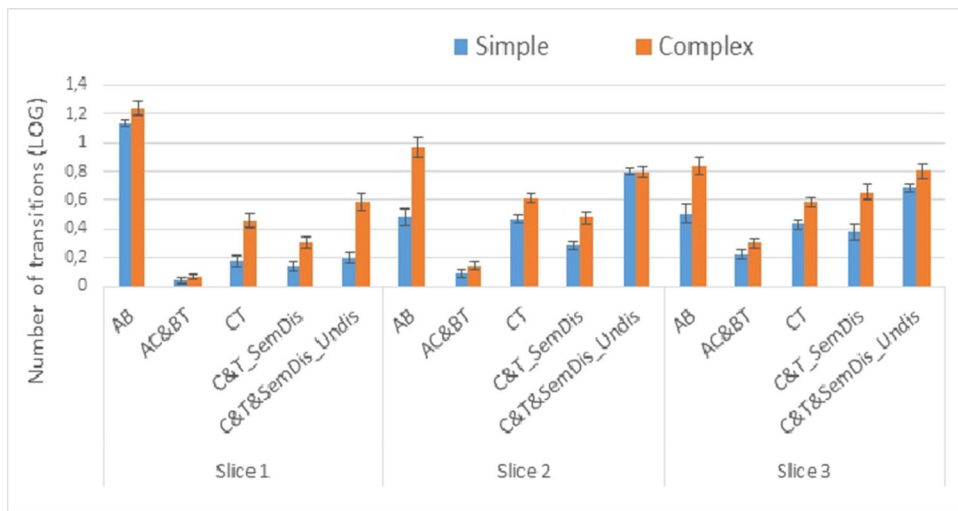
Fig. 3. Log of the mean number of each type of saccade in each slice in simple and complex trials (error bars are SEM).

AB transitions than all the other transition types for both Simple and Complex analogies ($p < .0001$). In Complex analogies, there were significantly more CT and, surprisingly, more C&T&SemDis_Undis, than AC&BT transitions, and more C&T&SemDis_Undis than C&T_SemDis ($p < .0001$). The last two results show that transitions involving the unrelated distractors were more numerous than transitions involving stimuli that were key to solving the analogy in complex trials. Again, this suggests that participants systematically compared unrelated distractors to C and to related items in the complex trials.

In S*lice2*, there was a significantly higher rate of AB transitions in complex than in simple problems. Within-condition analyses in the simple trials revealed higher rates of AB than AC&BT and C&T_SemDis ($p < .001$), and surprisingly, more C&T&SemDis_Undis transitions than all other types of transitions ($p < .0001$), and more CT than AC&BT. Thus, again AC&BT transitions were lower than the others ($p < .0001$). In the complex trials, there were significantly higher rates of AB transitions compared to other transition types, and fewer AC&BT than all the other types, and more C&T&SemDis_Undis than CTSemDis and CT transitions. Again, C&T&SemDis_Undis were at a high level in both conditions.

In S*lice3*, comparing complex and simple trials revealed significantly higher rates of AB and C&T_SemDis ($p < .0001$) in the complex trials which meant that, because finding a solution is difficult, participants continued to explore, check, and/or rerepresent the relation between A and B even at the end of the trial. Within-category comparison trials showed that, in simple trials, there were more C&T&SemDis_Undis than AC&BT, CT and C&T_SemDis, that is, transitions involving the solution set ($p < .0001$), and more AB than AC&BT transitions. In complex trials, there were higher rates of AB transitions than AC&BT and CT ($p < .0001$), the latter is compatible with the idea that participants needed to come back to A&B more often in order to assess the consistency of their answers. There were also more

C&T&SemDis_Undis than AC&BT and, importantly, CT ($p < .0001$), and fewer AC&BT transitions than all the other types ($p < .0001$).

The Tukey HSD post-hoc analysis also tells us, when we compare the first and third slices for simple versus complex problems, there were significantly more C&T&SemDis_Undis transitions in the third slice than in the first slice in both complexity conditions ($p < .0005$). This suggests that participants first concentrated on AB and applied the corresponding relation to CT, then, progressively checked other potential solutions, even though irrelevant, which is not predicted by the standard convergence views of analogical reasoning we reviewed in the Introduction. Tukey post-hoc analyses also revealed significantly more C&T_SemDis transitions in the third slice than in the first slice in both complexity conditions ($p < .005$), suggesting a later check of the semantic distractors in both complexity cases.

7. Intermediate summary

AB transitions initially dominated in both simple and complex trials. Progressively, however, participants studied the solution set together with C. This suggests that more comparisons of the relations were necessary in the complex case. Beyond that, importantly, there were virtually *no AC* or *BT* transitions in the three slices, which argues against a strict interpretation of alignment-first models.

A second result was that comparisons involving unrelated distractors remained frequent at the end of the trial in *both* difficulty conditions, and more frequent than the relevant C&T_SemDis transitions, although, interestingly, they were more frequent at the beginning of the complex trials, suggesting that participants were exploring these options during an entire complex trial. This is not predicted by the "parallel-terraced scan" view of analogical reasoning, nor other type of models we are aware of. The analysis of gazes confirmed these results, with a flatter profile between AOIs in the complex condition, suggesting a more balanced approach to the AOIs across time slices than in the simple condition.

7.1. Discriminating between Simple and Complex conditions: SVM+LOOCV

As mentioned in the Introduction, our purpose was to identify small sets of transitions that defined a particular search-space *exploration strategy*. By this, we mean that the numbers of each transition-type in these small sets of transitions defined a particular search strategy and our hope was that we would be able to predict whether the problem being solved was Complex or Simple, or whether the problem would be solved correctly or incorrectly, based on the observed strategy (French et al., 2017 used a similar methodology to predict whether adults and children were doing a problem or whether the problem would be answered correctly or erroneously). SVM allows us to find small subsets of transitions that allow us to predict with a high probability whether the problem being solved was complex or simple, or would be solved correctly or incorrectly.

We hypothesized that small subsets of transitions used during the problem were indicative of a search strategy and could be used to predict whether a participant was doing a Complex

or a Simple problem. We selected small subsets of transitions from the following set of 13 transitions: {AB, AC, BT, BC, CT, AT, CSemDis, TSemDis, A&B_UnDis, A&B_SemDis, CUnDis, TUnDis, and SemDisUnDis).

The only grouped transitions were A&B with SemDis, and A&B with UnDis. This is because A and B have the same status in the above pair. The same was done for ASemDis and BSemDis. In contrast with our behavioral analysis, we dissociated CSemDis, TSemDis, and transitions with UnDis, because the predictability of each transition type might differ. Also, SVM algorithms allow us to look at a relatively high number of transitions, and we took advantage of this.

We used the normalized number (by trial) of transitions as input to our SVM-LOOCV classifier. Note that relatively less frequent transitions may, in principle, contribute to distinguishing conditions if they appear in one condition and not in another condition and, on the other hand, frequent transitions in both conditions may not contribute to differentiate them.

We felt that small subsets of these transitions (three or fewer transitions) provide more search-strategy information than large subsets. For example, if we take a large enough set of transitions, the SVM algorithm would almost certainly have been able to discriminate Complex problems from Simple problems on that basis. But too large a number of transition-types tells us little about the search-strategy used by a person to solve the problem. Hence, we considered small subsets of three or fewer transitions that had the highest discriminative power.

We predicted that participants would have differing numbers of these transitions or subsets of transitions depending on the problem complexity and that these differences would predict whether the problem under consideration was Simple or Complex, or in a second analysis, whether it was an error or a correct trial.

In order to do this, we coupled the SVM with an LOOCV (Geisser, 1975; Miller, 1974; Stone, 1974; for a review, Arlot & Celisse, 2010), which, in the case of our analysis of transition profiles, worked as follows. We selected one of the N problems solved by the participants and set that problem aside. Then, for each of the remaining N–1 problems, which we designated as the SVM training set, we considered various sets of transition profiles such as {BT, TSemDis, CSemDis}. We then counted the number of each of these transitions made while solving the problem, *averaged* over all participants. We trained the SVM using these vectors for each problem in the SVM training set until it learned to correctly classify each of the N–1 problems in the training set as "Simple" or "Complex" (or as "error" or "correct") in our second analysis. We then gave the SVM the problem that it had not seen and saw if it was capable of classifying it correctly as either "Simple" or "Complex," based on how the other N–1 problems were classified. We applied this leave-one-out training-and-testing procedure to all of the problems. We applied this reasoning to the first time slice and the third time slice, corresponding to the beginning and the end of the search. We arbitrarily set at 0.75 the level of "good predictability" (classifying three trials out of four in the correct category, either Complex or Simple). When considering pairs or triplets of transitions, we kept only those transitions that increased the level of predictability of smaller subsets of transitions. For example, if AB alone had a predictive power of 0.75, we would consider only pairs of tran-

Table 1
Transition sets and individual transition frequencies for the first slice of Experiment 1

Slice 1 complex							
One transition		Two transitions			Three transitions		
Success rate	Transition set	Success rate	Transition set		Success rate	Transition set	
0.66	AB	0.9	AD/BD	AN/BN	0.81	AB	TD	DN
0.63	AN/BN	0.77	AN/BN	DN	0.81	AB	CN	DN
		0.75	AB	AD/BD	0.8	AT	AD/BD	TN
		Individual transition frequencies			0.78	AB	CT	DN
			AB	1	0.78	AB	TD	TN
			AD/BD	2	0.78	AB	AT	AD/BD
			AN/BN	2	0.78	CT	BC	TN
			DN	1	0.77	BT	CT	CD
					0.77	AT	AN/BN	TN
					0.75	AB	AC	TD
					0.75	AB	BT	DN
					0.75	AB	CD	TD
					0.75	AB	CD	DN
					0.75	AB	CN	TN
					Individual transition frequencies		
						AB	10
						DN	5
						TN	5
						TD	4
						CT	3
						AT	3
						CD	3
						CN	2
						BT	2
						AD/BD	2
						AN/BN	2
						BC	1
						AC	0

Note: D = SemDis; N = UnDis. AN/BN and AD/BD designate the average number of AN and BN transitions and AD and BD transitions, respectively.

sitions involving AB (e.g., AB CT) if their predictive power was higher than AB alone (e.g., 0.80).

For the Simple-Complex discrimination during *Slice1*, the SVM-LOOCV analysis gave the following prediction accuracy scale (Table 1). No single transition was above 0.75. AB and A_UnDis and B_Undis were at 0.66 and 0.63, respectively, confirming that transitions

involving A and B were (relatively) important at the beginning of the trial (AC and BT were at 0 and 0.13, respectively).

Prediction accuracy increased for sets with two transitions and reached our criterion for three pairs of transitions. {AB and A&B_SemDis} (0.75), {A&B_UnDis and SemDis_UnDis} (0.77), and {A&B_SemDis with A&B_UnDis} (0.90). This suggests that transitions involving A&B and distractors, both semantically related and unrelated, significantly improve prediction accuracy. Once again, this is compatible with the idea that the search space was broader from the start in the case of complex analogies.

When we added a third transition, we found that there were 14 triplets that produced correct Complex-Simple classification at 0.75 and above. In these triplets, AB was involved in a total of 10 of them. Interestingly, next on the list were five triplets of transitions containing SemDis_UnDis (DN in Table 1) and five subsets containing T_UnDis (TN in Table 1), confirming that, from the start, the Complex condition was characterized by a broader search space, that is, one involving more items, including semantically unrelated ones. This strongly suggests that the difference between Simple and Complex trials is a matter of finding the AB relation and realizing that the two types of distractors (UnDis and SemDis, T and N, respectively, in Table 1) are not the answer (T). At this stage, AC_BT had no discriminative power.

For *Slice3*, the most discriminative single transitions were AB and CSemDis, at 0.67. Looking at the 11 pairs of transitions beyond 0.75 showed that AB (4) and CSemDis (8) were particularly predictive, meaning that CSemDis played a central role in reaching a solution at the end of the trial. Adding a third transition gave a total of 22 sets of transitions producing a discrimination accuracy above 0.75. Once again, AB was involved in 13 out of the 22 triplets. The other important transitions were CSemDis (14) and CUnDis (9). This result suggests that, by the end of the trial, Complex trials are characterized by transitions involving C on the one hand, and SemDis and UnDis, on the other hand. This confirms and extends previous results (French et al., 2017; Thibaut, French, Missault, Gérard, & Glady, 2011)—namely, that complex problems involve more comparisons between C and both types of distractors, and continued focus on AB transitions. This, together with seven A&B_UnDis transitions (shown as AN/BN in Table 2) confirms that participants saccaded to unrelated distractors more in the complex problems, suggesting that they found it difficult right up to the end of the trial to decide whether unrelated distractors were a solution. The presence of AB, of transitions between AB and other AOIs at the end of the trial, is compatible with the idea that returning to the AB transition occurs more frequently in the Complex case, thereafter comparing C with the Target, SemDis, and UnDis. As far as we know, this is not predicted by any current model of analogy-making. Current models predict that, as participants get closer to answer selection, the number of saccades to semantically unrelated distractors should fall essentially to zero since they have, presumably, made their decision and will no longer need to saccade to semantically unrelated items. This is clearly not the case.

This confirms the previous (behavioral) analysis and suggests that, for complex problems, which require the space of possible solutions to be explored more thoroughly, participants look at *all potential solutions, including the unrelated distractors throughout the course of the problem.*

Table 2
Transition sets and individual transition frequencies for the third slice of Experiment 1

One transition		Two transitions			Three transitions			
Success rate	Transition set	Success rate	Transition set		Success rate	Transition set		
0.67	AB	0.88	BT	CD	0.94	BT	CD	CN
0.67	CD	0.81	CD	DN	0.88	AB	CT	CD
		0.8	AB	CD	0.87	AB	CD	CN
		0.77	AB	BC	0.84	AB	CT	AN/BN
		0.77	CT	CD	0.84	CD	CN	DN
		0.77	CD	BC	0.84	AB	AC	CD
		0.77	CD	AT	0.83	AB	CD	AN/BN
		0.77	CD	AD/BD	0.81	AB	BT	CT
		0.75	AB	BT	0.81	AB	CT	CN
		0.75	AB	CT	0.81	CD	CN	TN
		0.75	CD	TN	0.81	CD	BC	AD/BD
		Individual transition frequencies			0.78	AB	AC	BT
		AB	4		0.78	AB	AC	CT
		CD	8		0.78	AB	BT	TD
		CT	2		0.78	AB	BT	AN/BN
		BT	2		0.78	AB	BT	CN
		BC	2		0.78	CD	TD	AN/BN
		CT	2		0.77	CD	AD/BD	CN
		AD/BD	1		0.77	AC	CD	AN/BN
		AC	0		0.77	CD	AN/BN	CN
					0.75	AC	CD	TD
					0.75	CD	TD	CN
					Individual transition frequencies			
					CD	14		
					AB	13		
					CN	9		
					AN/BN	7		
					BT	6		
					AC	5		
					TD	4		
					CT	4		
					AD/BD	2		
					DN	1		
					TN	1		
					BC	1		
					AT	0		

Note: D = SemDis; N = UnDis. AN/BN and AD/BD designate the average number of AN and BN transitions and AD and BD transitions, respectively.

Even though we do not provide the behavioral analyses, we ran an identical SVM+LOOCV analysis to see which strategies led to correct or incorrect answers in the two groups of problems, to parallel the ones provided by Thibaut and French (2016) and French et al. (2017). Indeed, finding that these two types of conditions had different profiles in adults would be interesting, as French et al. (2017) focused on children only. The idea is to find out whether correct answers have their own signature, compared to errors, and whether an error profile can be detected from the start.

In *Slice1*, CUnDis was the best predictor (0.63) followed by CSemDis (0.62) and CT (0.60) and A&B_SemDis (0.60), showing that transitions involving both distractors contributed to distinguish Error trials from Correct trials (all Complex since there were no errors in the Simple condition). No pair reached our expected level of 75%, and only one triplet reached it, AT, BT, and A&B_UnDis (0.80), that is, pairs involving A and B with the target and the distractors. Thus, relating A and B to the solution and comparing them with unrelated distractors was a crucial factor distinguishing the two types of trials, early on in the trial.

In *Slice3*, CSemDis was a good unique predictor (78% of accuracy) followed by A&B_SemDis (0.65) and TUnDis (0.62). For pairs of transitions, CSemDis appeared nine times out of 15 pairs, confirming that controlling SemDis is a major feature of correct answers. All other transitions were evenly distributed and less frequent (three or less). Among the 31 triplets beyond 0.75, 29 included AB (AB, A&B_SemDis, and A&B_UnDis), and 38 transitions contained SemDis. This, again, argues in favor of the importance of a correct encoding of AB and careful control of SemDis against other options. Importantly, it should be noted that there were 13 transitions containing UnDis were also present, confirming the role of UnDis in taking a decision. Thus, by the end of the trial, including AB and SemDis in the decisional process seems to be a major feature distinguishing correct from error trials.

This result parallels what we observed in the prediction of whether a problem was Simple or Complex. Distinguishing between correct and error trials heavily relies on AB and on the exploration of transitions to semantically related distractors but, also, to unrelated distractors until the end of the problem. They also confirm what Thibaut and French (2016) found for children: errors and correct trials differ in their signature.

8. Discussion

Our study extends previous eye-tracking studies using analogies involving words rather than images. Bethell-Fox et al. (1984) used analogies defined around perceptual dimensions, whereas Gordon and Moser (2007), Thibaut and French (2016), Vendetti, Starr, Johnson, Modavi, and Bunge (2017; Starr et al., 2018) used pictures of objects or scenes.

Our data confirmed the projection-first strategy, that is, there were significantly higher rates of AB and CT saccades in participants' patterns of visual search compared to AC and BT saccades, for both Complex and Simple conditions. Thus, participants mostly infer the relation between the pictures in the A:B pair and apply it to the C-solution set (CT saccades). Neither Simple nor Complex problems elicit a significant amount of AC or BT saccades, that is, alignments, or any sign of relational priming as postulated by Leech et al. (2008).

The second purpose was to assess the impact of trial difficulty. In the second and third slices, participants tried multiple hypotheses in order to make sense of the analogies, or tried to rerepresent the relation between A and B after testing their initial hypotheses more often in the Complex condition than in the Simple one. This is consistent with the idea (Bethell-Fox et al., 1984) that at this stage no response has yet been eliminated, not even the nonsemantically related distractors (UnDis). The lower proportion of CT transitions in complex trials makes sense since participants are checking *all of the items* to make sure that, indeed, T is the correct solution. By contrast, for simple trials, response uncertainty is low and it is, therefore, not necessary to check the Unrelated Distractors.

However, the most unexpected result of the present work was the significant number of transitions toward unrelated distractors, together with semantic distractors, which, as shown by the SVM analyses, also had a high discriminating power when comparing Simple and Complex problems. In the Complex condition, the number of these transitions was greater than the mean number of CT transitions alone or the mean number of AC&BT transitions, or even the mean number of transitions from C or T to the semantically related distractors. *This is predicted by none of the models of analogical reasoning that we are aware of*. This makes sense if one considers that deciding that something is unrelated is especially hard in a Complex case or when confidence is lacking in the correct target relation. These controls, however, were less expected for unrelated distractors at the end of the trial than at its beginning. In short, unrelated information should have been discarded earlier on. The analysis of the AOIs revealed a flatter gaze profile in the complex condition across time slices, showing that participants distributed their looking times evenly across stimulus types when trials were more difficult.

Our main conclusion is that the task difficulty influenced the time course of the trial. Even though Complex and Simple trials resemble one another (e.g., same AB transitions at the beginning), they also differed in specific ways. In general, the time course of our verbal analogies was similar to previous results (e.g., French et al., 2017; Thibaut & French, 2016). Complex trials generated more exploration of the distractors and of the A-B pair, particularly at the end of the trial, which was unexpected.

8.1. Experiment 2: Scene analogies. Comparison between two-relation and single-relation problems

The present experiment is an extension of the Scene analogy used by Richland et al. (2006). We compared two types of problems, one with two relations (Complex) and the other with a single relation (Simple). As in Experiment 1, the main focus is on problem complexity, which can induce novel strategies or at least can lead to search-strategy adaptations in different contexts. This implies focusing on certain questions, such as when AC and BT alignments take place, whether distracting or irrelevant information is looked at and rapidly discarded or whether it is processed throughout the trial, and so on. It can be argued that for more complex problems, it is harder to establish the items that play analogous roles in order to arrive at a solution. Our hypotheses are similar to the ones in Experiment 1.
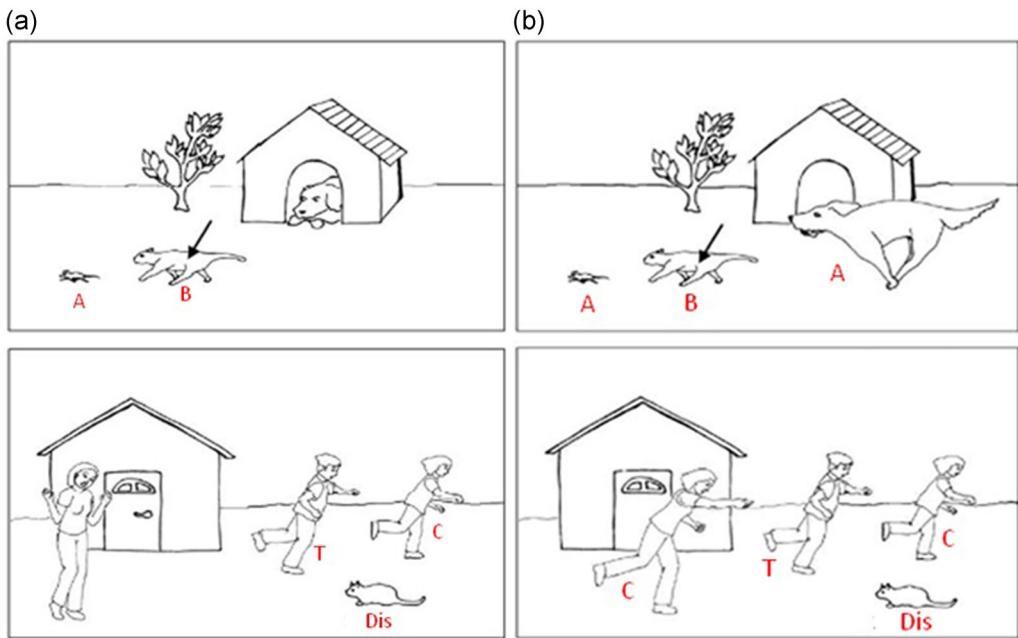
Fig. 4. (a): Simple scene analogy with a single salient relation in the base domain, in the case "a cat chasing a mouse." (b). Complex scene analogy with two salient relations in the base domain: "a cat chasing a mouse" and "a dog chasing a cat," where B is the targeted stimulus and T is the participants' choice.

8.2. Participants

Participants were 25 students at the University of Burgundy ($M = 21.6$ years; $SD = 2.2$; range = 19–26 years). They participated voluntarily or for course credit and were unaware of the experimental rationale. All participants had a normal or corrected-to-normal vision. In the latter case, it was checked that glasses did not interfere with data collection.

8.3. Materials

The task consisted of 14 trials (two training trials and 12 test trials, six Simple problems and six Complex problems). The scenes were the same as those used in Richland et al. (2006). Their list of stimuli was slightly adapted for the present experiment. Here, we used only the "distractor" condition used in Richland et al., which means that we did not have a "no-distractor" condition. All trials were composed of two scenes, a base scene in the upper panel of the figure and a target scene in the lower panel (Fig. 4). A distractor was chosen from the base scene and, in a slightly modified form, was added to the target scene. This distractor in the target scene was both visually and semantically related to one item in the relation in the base picture. For example, there was a cat in the base scene (i.e., a cat is chasing a mouse) and there was also a cat in the foreground of the target scene that depicted "a boy chasing a girl." There were two levels of complexity. The first we called "simple" and consisted of a single

relation (Fig. 4a) and the second one was called "complex" and consisted of two relations (Fig. 4b). In the six simple problems, both scenes depicted a single interaction between two entities (e.g., a cat and a mouse), whereas, in the six complex problems, both scenes depicted an interaction between three entities (defining two relations, e.g., a dog chasing a cat that was chasing a mouse, see Fig. 4). Participants had to determine the item in the lower drawing that best corresponded to the item in the upper drawing that was indicated with an arrow. The order of presentation of the test trials was random.

Each trial consisted of two scenes (501×376 pixels for each scene) each containing either five black-and-white (BW) line drawings framed by a black rectangle in the single-relation problems, or six BW line drawings for two-relation problems. The scenes were displayed on a Tobii T120 eye-tracker (120 Hz) with 1024×768 screen resolution using an E-Prime$^©$ software (version 2.8.0.22) embedded in a Tobii Studio (version 2.1.12) procedure to record participants' gazes.

These stimuli were labeled (for the purposes of data analysis, not in the experiment itself) A, B, C, T(arget), and Dis (the Distractor that was perceptually and semantically similar to the object designated with an arrow in the upper scene). In the single-relation condition case, the mouse and the cat were A and B, respectively, as shown in Fig. 4a, and the dog played no role in the targeted "chasing" relation. In the lower scene, the woman was standing still on the left, playing no role in the relation between the boy and the girl. In the two-relation scene, two stimuli played the role of A in Fig. 4b—namely, the mouse and the dog—and two stimuli played the role of C (the woman and the girl). The cat in the foreground of both Figs. 4a and b was the distractor (Dis).

Fig. 4a depicts a one relation—simple—problem and Fig. 4b depicts a two-relation—complex—problem, with the base scene in the upper panel and the target scene in the lower panel. The arrow in the upper scene points to a stimulus (the cat) and the participants must find the relational equivalent of the cat in the lower scene. By convention (see text), the non-targeted objects in the upper scene were called A (the dog and the mouse) and the designated object was called B. In the scene in the lower panel, C designated the woman and the girl and the target T was the boy. The cat in the lower panel is the distractor (Dis). Note the differences with the Simple condition (Fig. 4a) in which the stimuli are the same, with the exception that neither the dog nor the woman is participating in the action (the dog is in the doghouse and the woman is standing still to the left of the scene).

8.4. Procedure

Test sessions took place in a quiet experimental room in our laboratory. Each participant was tested individually. The distance between the participants' face and the screen was approximately 70 cm. After the eye-tracker was calibrated, participants were tested in the Scene analogical reasoning task. Participants were first shown a practice trial. When they had given an answer, the experimenter asked them to justify their answer and provided feedback. In the event of an incorrect justification, the trial was explained in terms of the relations linking A and B on one side, and C and T on the other side. For the test trials, participants received no further feedback or information about whether they had replied correctly

or not. Eye-tracking data were recorded when the presentation of the problem started and stopped when an answer was given.

9. Results

Before analyzing the time course of eye movements, we checked whether Complex trials were, in reality, more difficult than Simple trials. The mean number of correct answers was significantly lower for Complex than for Simple scenes: $t(24) = 2.7$, $p = .015$; $\eta^2_p = 0.22$, with $M = 83.3\%$ and 73.3% correct for Simple and Complex scenes, respectively. RT analyses showed responses to Complex scenes were significantly slower than for Simple scenes: $t(24) = 7.87$, $p < .01$, $\eta^2_p = 0.25$, with $M = 4775$ and 5925 ms for Simple and Complex scenes, respectively. Thus, the two conditions do, indeed, differ in difficulty, which, as intended, raises the question of differences in search strategies.

9.1. Eye-movement analysis

No trials were rejected because of insufficient gaze time (i.e., more than 50% of the gaze time was not recorded). As in the first experiment, the dependent variables were the percentage of total looking time and the number of saccades. As in Experiment 1, we first analyzed fixation durations (gazes) on each stimulus type (AOI) and, second, in order to compare the Single-relation and Two-relation trials, we analyzed the number of transitions and focused on the distribution of key saccades throughout the trial. As before, we divided all trials into three equal time slices in order to capture differences between the two conditions temporal dynamics.

9.1.1. Gaze (AOI) analysis

As in Experiment 1, we performed an analysis on fixation durations on five AOIs (A, B, C, T, and Dis, see below). We defined the AOIs as A for the nontargeted item of the relation (e.g., in Fig. 4a and b, A is the mouse), and B for the targeted item (the stimulus pointed to by the arrow). In the two-relation (complex) case, we computed the mean for the two nontargeted objects under A (e.g., in the example above, the mouse and the dog). In the below scene, C was the nonrelational solution involved in the relation (e.g., the girl in Fig. 4a). In the two-relation case, C was the aggregation of the two nontargeted stimuli in the relation (e.g., the woman and the girl in Fig. 4b). T(arget) was the correct relational answer. The distractor (Dis) was the same stimulus as stimulus B from the above scene (i.e., the cat). This distractor (Dis) is an important stimulus to be checked since it was meant to be perceptually and semantically related to the designated stimulus in the base (i.e., upper) scene and, thus, to attract participants' attention in the target (lower) scene. Note that it is a distractor because it is perceptually and semantically related to the equivalent stimulus in the base scene. For each participant, the time spent on a given AOI (e.g., A) was defined as the mean proportion of looking time spent on this AOI across the six trials defining a condition.
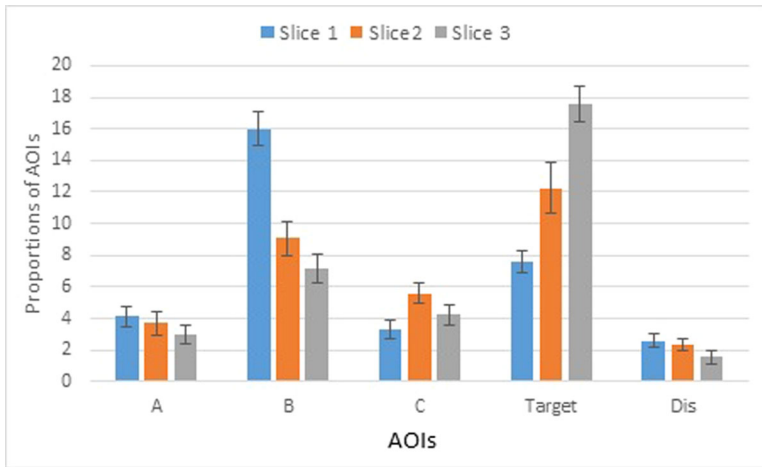
Fig. 5. Time slice by AOI interaction. It shows that participants were gazing at B in the first slice and also to the target. This pattern was reversed at the end of the trial. A, C, and distractors were less frequently focused on.

9.1.1.1. Fixations: We focused on the time course of fixations toward A and B compared to C and T, and on the difference between Complex and Simple trials in each time slice. A three-way repeated measure ANOVA, with Type of Stimulus (A, B, C, T, and Dis), Complexity (One-relation and Two-relation), and Slice (first, middle, and last) as within-subject factors, was performed on the mean percentage of gazes for the five AOIs in order to assess the temporal dynamics of rates of fixations. The analysis revealed a main effect of AOI, $F(4, 92) = 122.3$, $p < .0001$, $\eta^2_p = 0.84$; a main effect of Complexity, $F(1, 23) = 3918$, $p < .0001$, $\eta^2_p = 0.99$ interactions between AOI and Slice, $F(8, 184) = 30.63$, $p < .0001$, $\eta^2_p = 0.57$, and between AOI and Complexity, $F(4, 92) = 3.70$, $p < .01$, $\eta^2_p = 0.14$, which was the most interesting result. The triple interaction was not significant ($p > .1$). As Fig. 5 shows, for the AOI × slice interaction, participants mostly gazed at B at the beginning of the trial with fewer gazes at Target, followed by the reverse pattern later on. Note, interestingly, that A and C received fewer gazes. This was confirmed by Tukey HSD comparisons, showing that there were more B in slice 1 than in slices 2 and 3, more T in slice 3 than in slices 1 and 2 and in slice 2 than in slice 1. As for the intra slice pattern, it was similar to the pattern observed in the following interaction and we will not repeat it.

As for the Complexity × AOI interaction, Fig. 6 shows that both conditions had a similar gaze profile. B and T received a majority of gazes, and all the others far fewer. This was confirmed by an a posteriori Tukey HSD which showed that there were significantly fewer gazes toward A, C, and Dis in both the simple and the complex conditions and significantly more looks toward the target than all the other stimuli in both conditions. In the complex condition, there were significantly fewer gazes toward the distractor than toward the other stimuli. The comparison between the complex and the simple conditions proved nonsignificant for the five stimulus types. In sum, gazes revealed that B and T overwhelmingly dominated attention with a smooth transition from B to T from the first to the third slice, with C receiving intermedi-
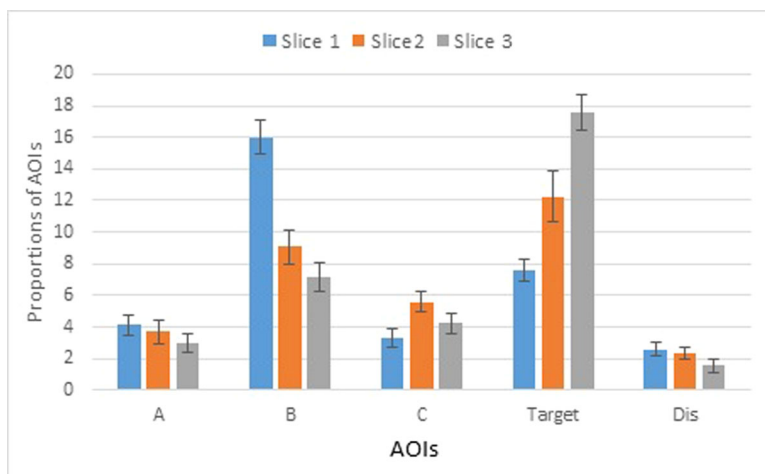
Fig. 6. AO× complexity interaction.

ate attention (recall that the scenes came from a corpus initially targeted at children, thus of moderate difficulty).

9.1.2. Transitions

In this analysis, we were primarily interested in sets of transitions that revealed strategy differences (if any) in solving Simple versus Complex scene problems. Six transition types were considered—namely: AB, CT, AC&BT, and C&T_Dis were also used in Experiment 1, with AC&BT being defined as the average number of AC and BT transitions. These two transitions were predicted by an alignment-first model because they involve alignments between equivalent stimuli in the two scenes. We also introduced two new types of transition, AT&BC, which was the average of AT and BC transitions. It plays the role of a control transition because, according to existing models, it should have no central role in solving the analogy and, hence, should be less frequent than AC&BT transitions, which are key transitions in some of the models. A second new type of transition was BDis. It refers to transitions between B and Dis. Richland et al. (2006) have shown that they are important because they refer to transitions between B, the stimulus pointed to in the base-above scene (e.g., a cat), and the equivalent stimulus in the target scene (e.g., another similar cat) and Dis is the most common mistake children made in their experiment. C&T_Dis is the average of transitions involving C and T toward the semantic-perceptual distractor, Dis, the cat in the foreground, with T being the relational Target. A three-way repeated-measures ANOVA, with Transitions (AB, CT, AC&BT, AT&BC, BDis, and C&T_Dis), Complexity (One-relation and Two-relation), and Slice (first, middle, and last) as within-subject factors, was performed on the log of the number of transitions (not normally distributed). The analysis revealed a main effect of Complexity, $F(1, 23) = 13.72, p < .001, \eta^2_p = 0.37$; of transition-type, $F(5, 115) = 39.16, p < .0001, \eta^2_p = 0.63$. There were interactions between transitions and complexity, $F(5, 115) = 8.96, p <$
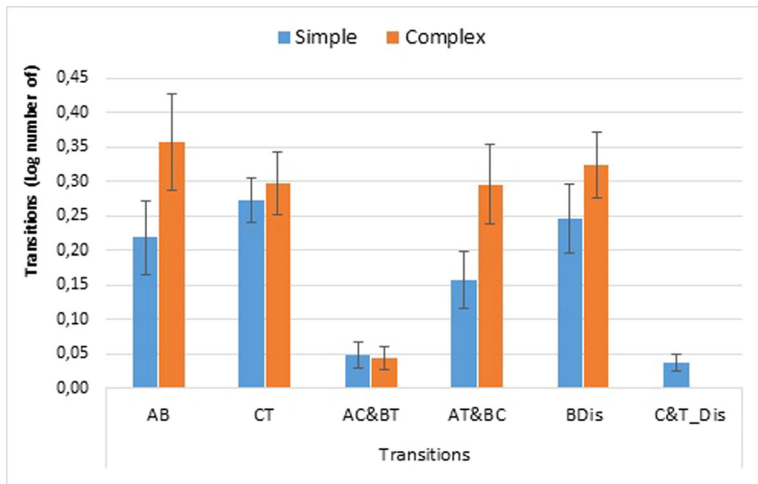
Fig. 7. Interaction between Complexity and Transitions. See text for transition names. (Error bars are SEM.)

.0001, $\eta^2_p = 0.28$ and between transitions and slice time, $F(10, 230) = 3.02$, $p < .005$, $\eta^2_p = 0.121$, the former interaction being the most interesting.

The Complexity × Transition-type interaction is interesting (Fig. 7) as it shows that there are more transitions for Complex scenes, for AB, for AT (i.e., A-Target), for BDis, and for CT. Once again, we found no evidence of AC or BT alignments. This is particularly interesting because there is a considerable comparison of A and the Target, but participants did not compare B with T, which is unexpected. BDis transitions make sense because B is the cat in the upper scene and semantically similar "cat" (Dis) in the lower scene. These are presumably perceived as the "same" stimuli.

The typical transitions, again, seem to be AB and then CT. The surprising feature is that the passage from the upper scene to the lower scene occurs via a transition from A (in the upper scene) to Target (in the lower scene). This was observed in both Complex and Simple conditions, even though there were more AT transitions in the two-relation Complex condition. Appendix A provides the confidence intervals corresponding to this interaction.

This was confirmed by a Tukey HSD post-hoc. As in Experiment 1, we only retained p values under .005. Results showed that there were significantly more AB and AT&BC in the Complex condition compared to the Simple condition ($p < .0005$).

Comparisons within conditions showed, in the Simple condition, that there were significantly more AB, CT, and BDis transitions than AC&BT, AT&BC, and C&T_Dis transitions, and surprisingly, more AT&BC than AC&BT ($p < .005$). In the Complex condition, there were significantly more AB and CT transitions than AC&BT and C&T_Dis, more AT&BC, and BDis, than AC&BT and C&T_Dis ($p < .005$). Overall, Simple scenes required fewer overall comparisons to establish the relationship between A and B, as being the key to solving the problem. Surprisingly, the two scenes aligned along nonanalogous stimuli—namely, A with T and B with C. Both conditions were organized around AB, CT, and BDis transitions,
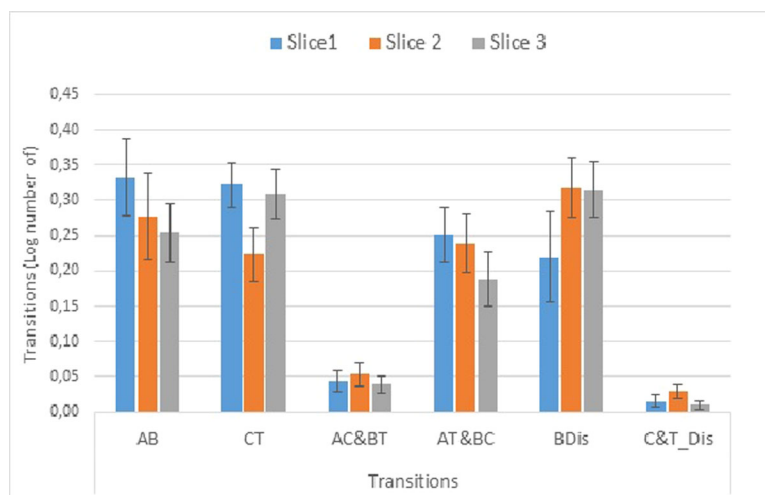
Fig. 8. Interaction between Transitions and Slice. (Error bars are SEM.)

overall. As in the first experiment, the very small number of AC and BT transitions does not fit with the predictions of alignment-first models, and the presence of AT&BC transitions is not predicted by any model we are aware of.

The Slice × Transition Interaction (Fig. 8) shows that the relevant transitions are produced during all three slices, except that the order slightly differs from one transition to another. A Tukey HSD post-hoc analysis showed, for slice 1, there were significantly more AB and CT transitions than AC&BT, C&T_Dis, and fewer AB than BDis ($p < .005$). Importantly, as mentioned above, there were few AC and BT transitions, which is difficult for an alignment-first model to explain, and few CDis and TDis transitions, which could be interpreted as meaning that participants saccaded only rarely to the Distractor item within the target scene. In contrast, BDis transitions show that the targeted item in the base and its perceptual/semantic counterpart were extensively compared.

Fig. 8 shows that the second and third time slices had similar patterns of transitions, with significantly less AC&BT and CD-TDis than all the others and less CT than BDis. For Slice 3, there were also significantly less AT&BC than BDis and than CT ($p < .005$). Overall, participants compared AB and CT, which were dominant across slices together with BDis, the latter becoming somewhat less important at the beginning of the trial (i.e., no significant difference with CT in the second and third slices, but in the first one). This is surprising since, given the dominant theories, the distractor should be discarded soon after the beginning of the trial, whereas data suggest that participants continue to saccade to it across the two scenes at the end of the trial. For all time slices, essentially no AC or BT transitions were observed.

9.2. Classification prediction based on subsets of transitions

We used the same SVM+LOOCV methodology as in Experiment 1. We started with a broad set of transitions in order to find out transitions, or pairs or triplets of transitions that

would differentiate Simple from Complex scenes. Again, in comparison to the behavioral analysis characterizing children's focus with a subset of main transitions as a function of the time slice, SVM analyses also give combinations of transitions with predictive power, not necessarily the most frequent ones. The set of nine transitions was AB, AC, BC, AT, BT, CT, BDis, CDis, and TDis. One purpose is to look at transitions involving Dis (the semantic-perceptual distractor in the below scene). In slice1, BDis was the most predictive transition (0.58) not at a high level, though. Three pairs were beyond 0.75, all involving Dis (CDis and TDis) and two CT. As for the two triplets, they confirm the presence of Dis in CDis (twice) and TDis. Thus, early control of Dis (with C and T) in the solution (target scene) space seems to be important. In Slice 3, no single transition approached our fixed threshold. A pair composed of BDis and CT, and two triplets both with BDis (2) CT (2) confirmed the predictive power of transitions involving BDis up to the end of the trial. The suggested pattern is that, by the end of the trial, participants compared C and T in order to decide which of the two is the correct solution. However, a parallel and last check that Dis is not the solution confirms that participants go for the distractors until they reach a decision. AC and BT had very low discriminative power alone (<0.30) and were not found in highly discriminative pairs or triplets.

10. Discussion

Experiment 2 sought to establish search profiles for two levels of relational Complexity in Scene analogies, and to ascertain whether participants would adapt their search strategies to the level of difficulty of the problems. Unsurprisingly, there was a main effect of the number of transitions with more transitions for Complex than for Simple problems. We observed relatively more AB transitions decreasing over time, while CT transitions increased and, again, we found no evidence of AC and BT alignment, even in the Complex condition. The transitions between B (in the base scene) and the Distractor (Dis, in the target scene) make sense since the two stimuli are the "same" (i.e., cats) in both scenes. This indicates that participants check the status of these two occurrences of the same stimulus in both scenes. The unexpected number of AT and BC transitions seems to suggest that participants also check both the relation B with respect to C and A with respect to T, presumably to be sure that these relations are not part of the solution. Complexity played a role, with the increased number of AB and AT&BC transitions for Complex problems. This suggests that establishing the role of B took required more saccades than in Simple problems and, in addition, required checking irrelevant pairings, such as AT, BC, and BDis. To summarize, complexity did not produce more alignments, in particular, AC and BT, but did require more comparisons in order for participants to establish the role of the designated stimulus (B), and to check irrelevant associations, such as AT and BC, while discarding Dis as an option over the course of the entire trial. This confirms the results of Experiment 1 in which distractors were more focused on when Complex problems were being solved. The SVM analyses confirm that transitions involving Dis played an important role in differentiating the two conditions.

10.1. General discussion

The overarching goal of the present paper is to better understand the impact of different analogy formats and complexity on analogy problem-solving strategies and find the gaze signatures characterizing each condition. We showed, using techniques from machine learning, that certain combinations of transitions, which define the search strategies employed by participants, were highly predictable of the type and difficulty of the analogies being solved.

Most of the available research using eye-tracking to study analogy-making has dealt with well-identified dimensions of analogy problems, such as those specified in matrices (Bethell-Fox et al., 1984; Chen et al., 2016; Hayes et al., 2011), and were restricted to only a single test format, such as scene analogies (Gordon & Moser, 2007) or proportional analogies (French et al., 2017; Starr et al., 2018; Thibaut & French, 2016; Thibaut et al., 2011). These previous studies did not manipulate for difficulty level, or did not systematically study the temporal dynamics of solution reaching. such as constructive matching (Bethell-Fox et al., 1984), in which participants first generate relations found in the base domain and, subsequently, apply them to the target domain. For example, Starr et al. (2018) translated each of the two hypothetical strategies into an algorithm combining early gazes and transitions which was then applied to each trial and led to its categorization as supporting one strategy type or the other. As the algorithms aggregated gaze length and positions, and transitions in a single measure, it was not meant to capture the temporal dynamics we were looking for here. Hayes et al. (2011) and Hayes and Petrov (2016) elaborated metrics to capture strategy regularities in Raven Progressive Matrices (Raven & Court, 1998). In Hayes et al. (2011), the authors measured participants' within the matrix and between the matrix and options transitions (see Bethell-Fox et al., 1984) and the relative weight of these two strategies. However, they did not look at the temporal dynamics of their integration. Hayes and Petrov (2016) introduced a method combining verbal protocols and pupillary responses, in an attempt to disentangle what they called exploration (new hypotheses) and exploitation (pattern descriptions) during solving. Their main point was to study "explore" and "exploit" search behaviors through their pupil diameter correlates and their distribution as a function of problem difficulty and time in a trial. Their approach, however, targeted these two broad categories (explore and exploit) and was not meant to study how participants distributed their gazes on stimuli as a function of their status (e.g., distractor).

Compared to these prior contributions, our paper raised five main issues—namely:

- (i) the generality of search patterns across analogy formats and across levels of difficulty;
- (ii) the role of alignment across formats and difficulty levels;
- (iii) the unanticipated focus on distractors over the course of a trial and as a function of difficulty level;
- (iv) the time course of the rejection of the distractors; and finally,
- (v) the predictability of subsets of transitions and AOIs.

Our most general result was that all analogy formats and difficulty levels elicit similar global search patterns, largely characterized by a projection-first approach. Increasing diffi-

culty led to more gazes within the source domain—A and B, and of AB transitions. Another general result was that AC and BT transitions were essentially absent in all conditions or played no role, in conjunction, in differentiating the conditions. We initially hypothesized that Scene analogies and more complex conditions might elicit these alignments, but this turned out not to be the case. Another significant result was the discovery, throughout trials, that there is a systematic examination of the distractors, both those related and unrelated to the A item (Experiment 1). As far as we know, this examination throughout the entire trial of distractors is not predicted by any current theory of analogy-making. Finally, explorations involving the distractors (AOIs and transitions with SemDis and UnDis, or BDis in the second experiment) could actually *increase* at the end of the trials. This result is also not predicted by the models of analogical reasoning we are aware of that see analogy solving as a progressive convergence toward the correct solution, which would imply a decrease in the examination of distractor items at the end of the trial.

On the other hand, participants did adapt their search strategies to different analogy formats, semantic distances between items, and the number of relations involved. Overall, Scene analogies gave rise to more gazes toward B than to A. Scene and Proportional analogies differed in that the target was examined earlier in Scene analogies, whereas Proportional analogy problems had more CT transitions than Scene analogy problems. The increased level of exploration of distractors (UnDis in Experiment 1 and Dis in Experiment 2) at the end of the trials was more pronounced in Complex conditions.

The examination of distractors, whether semantically related to C or not, throughout the trial is one of the main findings of the present research. Most models predict a progressive convergence toward an analogical match as the trial proceeds, with local, perceptual, irrelevant matches progressively discarded, and not semantically related items to C being the first to go. Experiment 1 revealed that, by the end of Complex problems, participants were increasingly testing both semantically related *and* unrelated distractors, suggesting that participants continued to test these solutions in parallel to the correct solution. Sometimes, the number of transitions to distractor items in Complex problems even exceeded the number of C-to-Target relations (CT).

How is this result to be interpreted? Our search-space hypothesis predicts that, when the relational solution is less salient, participants will tend to test unrelated distractors more systematically, whether or not they are relationally connected to C by a low saliency relation between A and B. This takes more time since an infinite number of descriptors is potentially available for each item (Goldstone, 1994b; Goodman, 1972; Murphy & Medin, 1985).

In addition, paradoxically, it can be easier to establish that a semantically related distractor is not the relational (i.e., analogical) solution because, once a relation between the distractor and C is identified, it is easy to check whether this relation holds in the base domain. Similarly, Experiment 2 showed that participants continued to compare B in the upper scene (the cat chasing the mouse) with Dis in the lower scene (the cat in the foreground, the semantic distractor) throughout the entire trial, rather than only in the early stages of solving the analogy (see also Gordon & Moser, 2007, for results and a discussion of the precedence of perceptual matches over relational matches). This suggests that there is no deactivation of the distractor over the course of the trial, which means that participants continue to transition to

it or gaze at it. This casts doubt on the standard view of convergence to a solution in which the items that are irrelevant to a solution are gradually discarded. At a more methodological level, our data suggest that segmenting trials into a number of time slices is an appropriate approach to study the evolution over time of solving an analogy problem.

Our *machine learning approach* provides a new tool aiming at better characterizing *the underlying dynamics of analogy-making*, which was developed and described in French et al. (2017). We looked for the smallest subset of transitions that could accurately predict as early in the trial as possible the type of problem being solved (Complex vs. Simple, Correct vs. Errors). Perhaps the most important contribution of the SVM+LOOCV technique is that these techniques show the extent to which, very early on in a trial, one can actually predict well above chance the difficulty of a problem based on the number and type of transitions observed, this corresponding to the search strategy adopted by the participant.

10.2. Modeling analogical reasoning

As already argued by Thibaut and French (2016), we believe that our data impose certain important constraints on models of analogy-making, not only in terms of what participants actually examine when solving a problem, but also *when* they examine various items and relations. Our experiments made no attempt to evaluate formal modeling approaches to analogical reasoning but, rather, examined some of the behavioral predictions derived from these models. The strengths and weaknesses of these models have been the focus of other papers (e.g., French, 2002; Gentner & Forbus, 2011, for reviews of computational modeling efforts).

10.3. What our data show

10.3.1 Relational-priming models

These models predict essentially none of the back-and-forth dynamics that we observed empirically as participants solve standard types of analogy problems (Leech et al., 2008). The underlying assumption is that once the relation between A and B is perceived, no further exploration is required and a solution to the problem is quasi-immediate. Our data clearly do not support this view of analogical reasoning for which the AB relation would prime the CTarget relation, once it has been discovered. Instead, we observed extensive evidence of systematic comparisons between C, the Target, and the distractors in both Experiments, including ongoing comparisons involving the unrelated distractors in Experiment 1. In the latter case, transitions between C and the Unrelated Distractors were among the most numerous transitions and, moreover, occurred throughout the trial, something that is certainly not predicted by an automatic-priming view that predicts a rapid selection of the relational solution, but no systematic gazes and comparisons with distractors at the end of the trial.

10.3.2 Projection-first models

Our results are largely consistent with the idea of an exploration of the source domain for relations, which are then generalized to the target domain. This is consistent with a model, such as LISA (Hummel & Holyoak, 1997) in which mappings are conceived as guided pattern-matching. For example, in the cat-mouse pair, pointing to the cat will activate the

proposition "*chase*(mouse, cat)." This should be followed by CT saccades corresponding to "*chase*(girl, boy)," since the relation *chase* exists in both domains. This approach predicts that the higher the activation of a relation, the lower the activation of other relations, which, over time, predicts less activation of items that are irrelevant to the solution of the analogy problem. It is, however, not clear how this model would account for the high level of transitions involving semantic distractors and, especially, transitions involving unrelated distractors throughout the trial, and for the TSemDis transitions at the end of the trials in Experiment 1.

10.3.3 Alignment-first models

These models are based on the alignment of entities that play equivalent roles. The basic idea is that participants start looking for mappings of various sorts, including perceptual mappings, which, at least initially, consist of many local mappings. The purpose is to discover mappings that transfer global relational meaning between the source and target domains. However, one of the main claims of the model is that there will be *between-domain alignments* (notably, AC and BT). However, for the two types of analogy formats presented in this paper, for each of the levels of difficulty of problems, and for each of the time slices within problems, we found no evidence of these cross-domain alignments. Further, there were numerous alignments not predicted by this model, in particular, transitions between AT and BC.

10.3.4. Parallel-terraced scan models

These models make no claim as to alignment-first or projection-first strategies, being rather a fluid combination of both approaches, based on activations in an associated semantic network. However, the lack of AC and BT transitions that we have shown empirically is problematic for this model, as is the initial overfocusing in children on relations centered around C (French et al., 2017; Thibaut & French, 2016). Also, they do not predict transitions to the distractors, especially the fact that these transitions tend to increase at the end of trials involving Complex trials.

10.4. Adapting models to the current results

Our paper offers a combination of three related results—namely, the generality of strategies across analogy-problem formats, the importance of search-strategy changes that depend on the intrinsic difficulty of the task, and the continued transitioning to distractors, including unrelated ones, throughout the trial.

Most approaches posit a type of cognitive-resource sharing, whereby the higher the activation of a particular relation, the lower the activation of other relations, which, over time, predicts less activation of items that are irrelevant to the solution of the analogy problem. Thus, one challenge for these models is to account for the high level of transitions involving semantic distractors in both experiments and, to an even larger extent, with unrelated distractors throughout the trial. These results would seem to suggest that the analogical choice is paralleled by a continued search for other, potentially less obvious solutions when the correct answer has a low level of activation. Continued saccading to unrelated distractors that have no clear semantic relation with C is problematic for most models.

Further, models of analogy need to move away from a clear separation of data representation and the processing of those representations, in the sense that processing (i.e., comparisons of pairs of stimuli) potentially leads to rerepresentation of the original data (i.e., possible relations between items in a pair). In the Complex problems presented here, we have seen how participants adapt their representations *as they are solving the problem*. A rerepresentation can also be conceived as dynamically changing built-in representations or as constructing novel representations on the fly. It is difficult to imagine that all possible relations necessary to solve any particular analogy problem could be built into the system a priori. These relations must be discovered on the fly and cannot reasonably be anticipated a priori (e.g., French, 1995; Hofstadter & Sander, 2013; Kokinov, Bliznashki, Kosev, & Hirstova, 2007; Schyns, Goldstone, & Thibaut, 1998). The issue we raise here is certainly not new. The possibility of preencoded versus dynamically encoded representations is a longstanding and difficult issue originating in early works on semantic memory (e.g., Smith, 1978).

10.5. Limitations of the present contribution

One limitation of the present work is its exhaustiveness. We have concentrated on types of analogy problems that are widely used in the research literature on analogy-making and for which the solution is to be found among a small set of options. Other types of problems are, of course, available. For example, one might argue that an analogy verification task, in which valid or invalid analogies have to be verified (e.g., "Is dog:bone::bird:seeds valid?"), would be a better case for alignments.

In conclusion, our hope is this paper will ultimately lead to appropriate adjustments to the current models of analogy-making that we believe have difficulty accounting for the results we have presented.

Acknowledgments

Open Research Badges

This article has earned Open Data and Open Materials badges. Data are available at https://osf.io/kp5j9/ and materials are available at https://osf.io/kp5j9/.

References

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*, 40–79.

Bendetowicz, D., Urbanski, M., Aichelburg, C., Levy, R., & Volle, E. (2017). Brain morphometry predicts individual creative potential and the ability to combine remote ideas. *Cortex*, *86*, 216–229.

Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence*, *8*(3), 205–238.

Bugaiska, A., & Thibaut, J.-P. (2015). Analogical reasoning and aging: The processing speed and inhibition hypothesis. *Aging, Neuropsychology, and Cognition*, *223*, 340–356.

Chen, Z., Honomichl, R., Kennedy, D., & Tan, E. (2016). Aiming to complete the matrix: Eye-movement analysis of processing strategies in children's relational thinking. *Developmental Psychology*, *526*, 867–878.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *826*, 407.

Doumas, L. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review*, *1151*, 1.

Duchowski, A. (2007). Eye tracking techniques. *Theory and practice* (2nd ed.). Springer.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, *411*, 1–63.

French, R. M. (1995). *The subtlety of sameness: A theory and computer model of analogy-making*. MIT Press.

French, R. M. (2002). The computational modeling of analogy-making. *Trends in Cognitive Sciences*, *65*, 200–205.

French, R. M., Glady, Y., & Thibaut, J. P. (2017). An evaluation of scanpath-comparison and machine-learning classification algorithms used to study the dynamics of analogy making. *Behavior Research Methods*, *494*, 1291–1302.

Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, *70350*, 320–328.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *72*, 155–170.

Gentner, D., & Forbus, K. D. (2011). Computational models of analogy. *Wiley Interdisciplinary Reviews: Cognitive Science*, *23*, 266–276.

Gentner, D., Holyoak, K. J., & Kokinov, B. (2001). *The analogical mind: Perspectives from cognitive science*. MIT Press.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *103*, 277–300.

Gick, & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*, 306–355.

Glady, Y., French, R. M., & Thibaut, J. P. (2017). Children's failure in analogical reasoning tasks: A problem of focus of attention and information integration? *Frontiers in Psychology*, *8*, 707.

Goldstone, R. L. (1994a). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *201*, 3.

Goldstone, R. L. (1994b). The role of similarity in categorization: Providing a groundwork. *Cognition*, *522*, 125–157.

Goodman, N. (1972). Seven strictures on similarity. In *Problems and projects*. Bobs-Merril.

Gordon, P. C., & Moser, S. (2007). Insight into analogies: Evidence from eye movements. *Visual Cognition*, *151*, 20–35.

Green, A. E. (2016). Creativity, within reason: Semantic distance and dynamic state creativity in relational thinking and reasoning. *Current Directions in Psychological Science*, *251*, 28–35.

Green, A. E., Kraemer, D. J. M., Fugelsang, J. A., Gray, J. R., & Dunbar, K. (2010). Connecting long distance: Semantic distance in analogical reasoning modulates frontopolar cortex activity. *Cerebral Cortex*, *201*, 70–76.

Green, A. E., Kraemer, D. J., Fugelsang, J. A., Gray, J. R., & Dunbar, K. (2012). Neural correlates of creativity in analogical reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(2), 264.

Hayes, T., Petrov, A., & Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on Raven's Advanced Progressive Matrices. *Journal of Vision*, *11*(10),1–11.

Hayes, T. R., & Petrov, A. A. (2016). Pupil diameter tracks the exploration–exploitation trade-off during analogical reasoning and explains individual differences in fluid intelligence. *Journal of Cognitive Neuroscience*, *282*, 308–318.

Hobeika, L., Diard-Detoeuf, C., Garcin, B., Levy, R., & Volle, E. (2016). General and specialized brain correlates for analogical reasoning: A meta-analysis of functional imaging studies: Meta-analysis of analogy brain networks. *Human Brain Mapping*, *375*, 1953–1969.

Hofstadter, D. R., & Sander, E. (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. Basic Books.

Holyoak, K. J. (2012). *Analogy and relational reasoning*. Oxford University Press.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *1043*, 427–466.

Keane, M. (1987). On retrieving analogues when solving problems. *Quarterly Journal of Experimental Psychology*, *391*, 29–41.

Kmiecik, M. J., Brisson, R. J., & Morrison, R. G. (2019). The time course of semantic and relational processing during verbal analogical reasoning. *Brain and Cognition*, *129*, 25–34.

Kokinov, B., Bliznashki, S., Kosev, S., & Hirstova, P. (2007). Analogical mapping and perception: Can mapping cause a re-representation of the target stimulus? *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Krawczyk, D. (2017). *Reasoning: The neuroscience of how we think*. Academic Press.

Leech, R., Mareschal, D., & Cooper, R. P. (2008). Analogy as relational priming: A developmental and computational perspective on the origins of a complex cognitive skill. *Behavioral and Brain Sciences*, *31*(04), 357–378.

Le Meur, O., & Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: Strengths and weaknesses. *Behavior Research Methods*, *451*, 251–266.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *254*, 431–467.

Miller, R. G. (1974). The jackknife–A review. *Biometrika*, *61*(1), 1–15.

Mitchell, M. (1993). *Analogy-making as perception: A computer model*. MIT Press.

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology*, *122*, 252–284.

Murphy, G. L. (2002). *The big book of concepts*. MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *923*, 289.

Rattermann, M. J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development*, *134*, 453–478.

Raven, J. C., & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales*. Oxford Pyschologists Press.

Rayner, K. (2012). *Eye movements and visual cognition: Scene perception and reading*. Springer Science & Business Media.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, *943*, 249–273.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *211*, 1–17.

Smith, E. E. (1978). Theories of semantic memory. *Handbook of learning and cognitive* processes.

Starr, A., Vendetti, M. S., & Bunge, S. A. (2018). Eye movements provide insight into individual differences in children's analogical reasoning strategies. *Acta Psychologica*, *186*, 18–26.

Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review*, *844*, 353–378.

Stevenson, C. E., Heiser, W. J., & Resing, W. C. (2013). Working memory as a moderator of training and transfer of analogical reasoning in children. *Contemporary Educational Psychology*, *383*, 159–169.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *291*, 41–78.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, *362*, 111–133.

Thibaut, J. P., French, R. M., Missault, A., Gérard, Y., & Glady, Y. (2011). In the eyes of the beholder: What eye-tracking reveals about analogy-making strategies in children and adults. *Proceedings of the Thirty-Third Annual Meeting of the Cognitive Science Society*, *33*, 453–458.

Thibaut, J. P., French, R. M., & Vezneva, M. (2010a). The development of analogy making in children: Cognitive load and executive functions. *Journal of Experimental Child Psychology*, *1061*, 1–19.

Thibaut, J. P., French, R. M., & Vezneva, M. (2010b). Cognitive load and semantic analogies: Searching semantic space. *Psychonomic Bulletin & Review*, *174*, 569–574.

Thibaut, J. P., & French, R. M. (2016). Analogical reasoning, control and executive functions: A developmental investigation with eye-tracking. *Cognitive Development*, *38*, 10–26.

Vapnik, V. (1999). *The nature of statistical learning theory*. Springer Science & Business Media.

Vendetti, M., Knowlton, B. J., & Holyoak, K. J. (2012). The impact of semantic distance and induced stress on analogical reasoning: A neurocomputational account. *Cognitive, Affective & Behavioral Neuroscience, 12*, 80–4–812.

Vendetti, M. S., Starr, A., Johnson, E. L., Modavi, K., & Bunge, S. A. (2017). Eye movements reveal optimal strategies for analogical reasoning. *Frontiers in Psychology*, *8*, 932.

Vendetti, M. S., Starr, A., Johnson, E. L., Modavi, K., & Bunge, S. A. (2017). Eye movements reveal optimal strategies for analogical reasoning. *Frontiers in Psychology*, *8*.

Vendetti, M. S., Wu, A., & Holyoak, K. J. (2014). Far-out thinking: Generating solutions to distant analogies promotes relational thinking. *Psychological Science*, *254*, 928–933.

Appendix A: Experiment 1. AOIs—confidence intervals
Simple Complex

		Mean	−95%	95%			Mean	−95%	95%
Slice 1	A&B	22.20	20.25	24.14	Slice 1	A&B	15.16	13.18	17.15
Slice 1	C&T	6.53	5.81	7.26	Slice 1	C&T	6.41	5.77	7.04
Slice 1	SemDis	2.52	1.52	3.51	Slice 1	SemDis	6.37	5.25	7.48
Slice 1	UnDis	2.09	1.19	2.99	Slice 1	UnDis	5.40	4.28	6.51
Slice 2	A&B	3.85	2.81	4.89	Slice 2	A&B	7.61	6.41	8.82
Slice 2	C&T	8.28	7.38	9.19	Slice 2	C&T	7.14	6.48	7.80
Slice 2	SemDis	10.88	9.71	12.05	Slice 2	SemDis	10.09	9.22	10.96
Slice 2	UnDis	10.51	9.54	11.48	Slice 2	UnDis	8.49	7.54	9.43
Slice 3	A&B	4.76	3.11	6.42	Slice 3	A&B	6.63	5.45	7.82
Slice 3	C&T	13.82	12.58	15.06	Slice 3	C&T	10.97	9.96	11.99
Slice 3	SemDis	7.59	6.22	8.95	Slice 3	SemDis	8.57	7.34	9.81
Slice 3	UnDis	7.16	5.82	8.50	Slice 3	UnDis	7.16	5.69	8.62

Appendix A: Experiment 1. Transitions—confidence intervals
Simple Complex

		Mean	−95%	95%		Mean	−95%	95%
Slice 1	AB	1.14	1.09	1.18	AB	1.24	1.14	1.34
Slice 1	AC&BT	0.04	−0.01	0.09	AC&BT	0.07	0.02	0.11
Slice 1	CT	0.17	0.10	0.25	CT	0.46	0.36	0.56
Slice 1	C&T_SemDis	0.14	0.07	0.21	C&T_SemDis	0.31	0.22	0.39
Slice 1	C&T&SemDis_Undis	0.20	0.12	0.28	C&T&SemDis_Undis	0.59	0.47	0.72
Slice 2	AB	0.48	0.35	0.61	AB	0.97	0.83	1.11
Slice 2	AC&BT	0.09	0.03	0.14	AC&BT	0.14	0.09	0.20
Slice 2	CT	0.47	0.41	0.52	CT	0.61	0.55	0.68
Slice 2	C&T_SemDis	0.28	0.22	0.34	C&T_SemDis	0.48	0.38	0.57
Slice 2	C&T&SemDis_Undis	0.80	0.75	0.85	C&T&SemDis_Undis	0.79	0.71	0.88
Slice 3	AB	0.51	0.36	0.65	AB	0.84	0.71	0.97
Slice 3	AC&BT	0.22	0.16	0.28	AC&BT	0.30	0.23	0.37
Slice 3	CT	0.43	0.37	0.50	CT	0.59	0.51	0.66
Slice 3	C&T_SemDis	0.38	0.26	0.50	C&T_SemDis	0.66	0.54	0.78
Slice 3	C&T&SemDis_Undis	0.69	0.63	0.74	C&T&SemDis_Undis	0.80	0.69	0.92

Appendix A: Experiment 2. AOIs—confidence intervals
Significant interaction involving complexity but no difference between simple and complex conditions.

Appendix A: Experiment 2. Transitions—confidence intervals

Simple Transitions	Mean	−95%	95%	Complex Transitions	Mean	−95%	95%
AB	0.22	0.11	0.33	AB	0.36	0.21	0.50
CT	0.27	0.20	0.34	CT	0.30	0.20	0.39
AC-BT	0.05	0.01	0.09	AC-BT	0.04	0.01	0.08
AT-BC	0.16	0.07	0.24	AT-BC	0.30	0.18	0.41
Bdis	0.25	0.14	0.35	Bdis	0.32	0.22	0.42
CDis-TDis	0.04	0.01	0.06	CDis-TDis	0.00	0.00	0.00