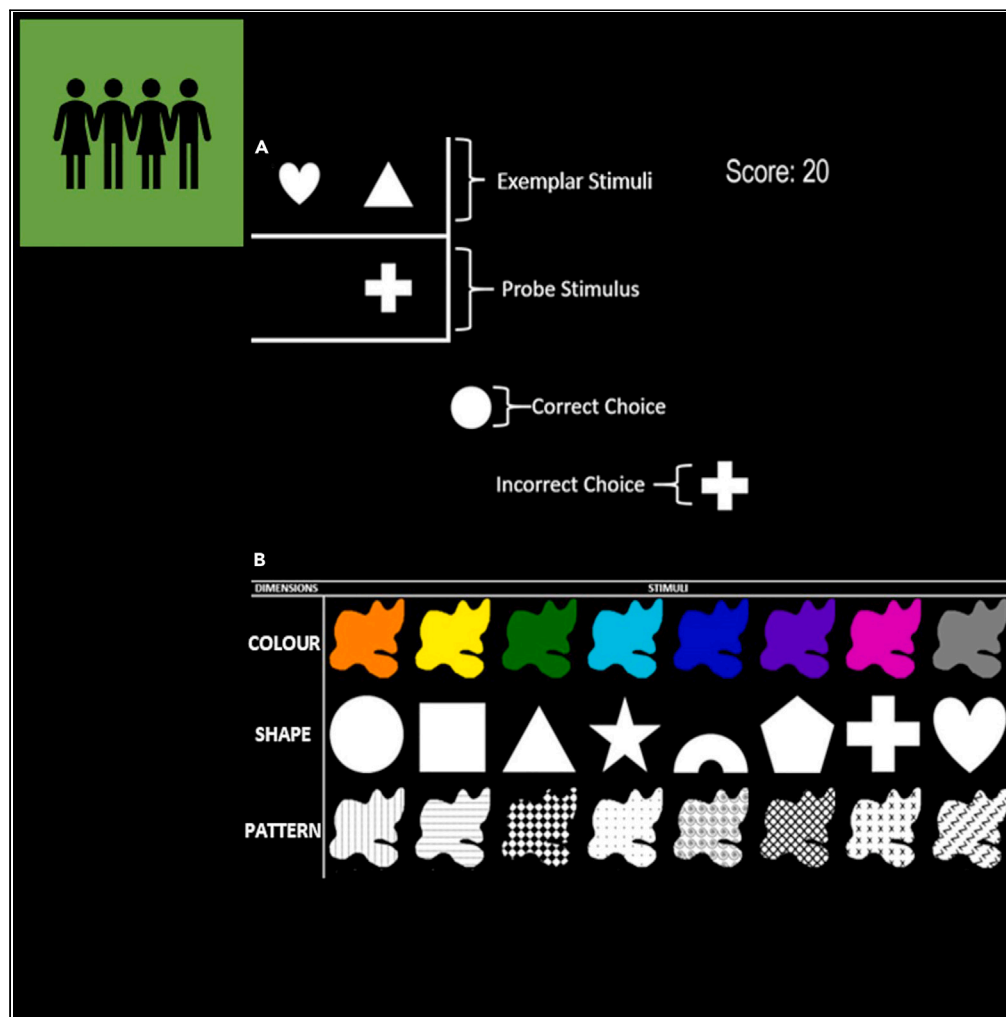


Article

Relational complexity influences analogical reasoning ability



Kevin Leonard,
Parisa Sepehri,
Breanna Cheri,
Debbie M. Kelly

debbie.kelly@umanitoba.ca

Highlights

A procedure for
examining analogical
reasoning that does not
require language

Our procedure permits
comparative studies of
analogical reasoning

For humans, the
complexity of an analogy
influences their ability to
reason

Article

Relational complexity influences
analogical reasoning abilityKevin Leonard,¹ Parisa Sepehri,¹ Breanna Cheri,¹ and Debbie M. Kelly^{1,2,*}

SUMMARY

Human language without analogy is like a zebra without stripes. The ability to understand analogies, or to engage in relational reasoning, has been argued to be an important distinction between the cognitive abilities of human and non-human animals. Current studies have failed to robustly show that animals can perform more complex, relational discriminations, in part because such tests rely on linguistic or symbolic experiences, and therefore are not suitable for evaluating analogical reasoning in animals. We report on a methodological approach allowing for direct comparisons of analogical reasoning ability across species. We show that human participants spontaneously make analogical discriminations with minimal verbal instructions, and that the ability to reason analogically is affected by analogical complexity. Furthermore, performance on our task correlated with participants' fluid intelligence scores. These results show the nuance of analogical reasoning abilities by humans, and provide a means of robustly comparing this capacity across species.

INTRODUCTION

Analogies are pervasive across human language and culture, defining the way we interact within, and understand, our world.^{1,2} When Forrest Gump told us that "... life was like a box of chocolates. You never know what you're gonna get",³ we intuitively understand that he was not suggesting a literal cocoa-quality to life. Instead, he was acknowledging the unpredictability of life that resembles the arbitrary nature of choosing one chocolate from an assortment of possibilities. Analogical reasoning, like that needed to understand Forrest's insight, involves the transfer of knowledge between seemingly disparate objects or ideas by finding the overlap or correspondence between them.⁴ It requires the ability to think abstractly, pushing past simple perceptual characteristics to assess relational and configural details. Although previously thought to be merely a particular example or type of cognition,⁵ the ability to make the relational assessments necessary to understand analogies is so fundamental to human cognition that it has led some researchers to posit that analogical reasoning is the core of cognition.⁶ Indeed, analogies are present in the way we understand language,⁷ social structures,⁸ and causal relations.⁹ Given that the use of analogy weaves through so many of the abilities that define human cognition, it is not surprising some researchers have argued that the capacity to understand higher-order relations communicated through analogy is one of the major gaps between human and non-human animal cognitive abilities.¹⁰ However, this apparent difference may be due to methodological differences in the way humans and animals are tested for their analogical reasoning ability.¹¹

Researchers use a variety of procedures to assess analogical reasoning by humans.¹² Many such tests involve the identification or application of verbal analogies, which make it difficult to construct equivalent tests for animals. Even within the realm of verbal analogies, there are distinctions between classic analogies of the form A:B::C:D (verbalized in the form "A is to B as C is to D," and also sometimes depicted as A:A'::B:B') and so-called "problem-solving" analogies.^{9,11} Classic verbal analogies place together pairs of words (e.g., HAND:FINGER::FOOT:TOE). When presented with such a word pairing, participants must determine whether an analogy is present. To solve the example analogy, participants must understand the words, infer the relationship between HAND and FINGER, map the relationship between HAND and FOOT, and apply that relationship to FOOT and TOE to assess whether an analogy is present.¹³ Problem-solving analogies, alternatively, require the application of knowledge or ideas from one domain to solve some target problem in a seemingly unrelated domain. Perhaps most famous of these problem-solving analogies is the Duncker's radiation problem.¹⁴ During this task, participants are typically asked to

¹University of Manitoba,
Winnipeg, MB R3T 2N2,
Canada

²Lead contact

*Correspondence:
debbie.kelly@umanitoba.ca
<https://doi.org/10.1016/j.isci.2023.106392>



pretend that they are a doctor with a patient who has an inoperable tumor. Participants must produce a solution that will destroy the tumor without damaging healthy tissue, namely the application of small amounts of radiation that converge on the tumor site. Crucially, if participants learned to solve an analogous problem within a different context (namely one involving military tanks sieging a castle), they were more likely to be able to solve the radiation problem, despite obvious differences in the surface content of the tasks.¹² The ability to reason in this manner, as a method of reducing uncertainty when dealing with complex problems, is a useful method for handling real-world scenarios.¹⁵ Although these verbal analogical reasoning tasks resemble the ways in which humans often use analogy in their everyday lives, they are particularly difficult to translate into something testable in a comparative context due to their reliance not only on language but also additionally on some fundamental real-world knowledge of concepts, such as tumors and tanks in the case of Duncker's problem. Thus, comparative researchers of analogical reasoning must find ways of presenting similar analogical concepts in ways that do not rely on linguistic or cultural knowledge.

In addition to verbal tests of analogy, analogies can be presented in a purely pictorial form, which lends itself better to comparative applications. These kinds of tests are used in developmental psychology, wherein children are required to complete analogies using concrete, real-world knowledge⁹ or abstract stimuli.¹⁶ A recent study presented adult participants with abstract pictorial stimuli in the form of line drawings.¹⁷ When pairs of stimuli were perceptually similar, participants could easily and quickly use perceptual cues to discriminate, but as the perceptual cue faded and the pairs of stimuli became more visually dissimilar, participants needed to transition to a conceptual approach. If the participants were also required to perform a concurrent working memory task, their performance dropped if they were in the "conceptual mode", indicating that the ability to make the conceptual-relational assessment relies on working memory.

Similarly, the Raven's Progressive Matrices, a popular test of fluid intelligence, uses abstract pictorial stimuli and is solved using strategies that are also relevant for solving analogies.¹⁸ The Raven's test is designed such that items progress in difficulty; easy items are presented early in the test (solvable by more than 90% of participants), whereas challenging items are presented at the end (solvable by fewer than 10% of participants¹⁹). There is theoretical and empirical support showing that the more difficult items of the Raven's Progressive Matrices are more taxing on and require greater involvement of working memory capacities.²⁰ For example, Little and colleagues¹⁹ showed an increase in item-wise correlations with measures of working memory capacity across item difficulty in both behavioral and simulated experiments. Furthermore, recent research has shown convincing evidence that fluid intelligence is anchored by attentional processes, with neurophysiological evidence demonstrating links between fluid intelligence and the brain's multiple-demand system.²¹ Due to the near ceiling performance on early items of the Raven's Progressive Matrices, not only are the correlations with working memory highest for items near the end of the test but also the correlations are near zero for the earlier items because, regardless of working memory abilities, most individuals are able to solve these problems.

Within the comparative literature, there is debate about the ability of animals to perform analogical reasoning, although there is a breadth of research findings supporting the existence of precursor faculties necessary for analogical reasoning.²² Analogical reasoning, also known as second-order relational reasoning due to its requirement for subjects to compare two sets of stimuli to evaluate whether the stimuli relationships within each set is the same, builds upon the precursor abilities of relatively simpler first-order same-different discrimination. This refers to the ability to determine if two individual stimuli are the same or different from one another, and similar to analogical reasoning, was once thought to be an exclusively human ability. However, mounting evidence of same-different (i.e., first order) discrimination abilities have accumulated for a variety of species.²³ This includes not only species noted for their advanced cognitive abilities such as non-human primates,²⁴ corvids,²⁵ and parrots²⁶ but also species traditionally viewed as having "lower cognitive" abilities such as pigeons,²⁷ rats,²⁸ and honeybees.²⁹ However, the capacity of animals to perform higher-order relational reasoning has been more elusive. This has been argued to be in part due to qualitative differences in the *same* and *different* concepts possessed by humans and animals, in that the human understanding of these ideas has specific linguistic qualities not present in animals.³⁰ Language-trained primates outperform language-naïve primates on these types of tasks, suggesting that having experience with concrete symbolic systems may be important for analogical reasoning.³¹ Yet, there is some evidence that primates without language training can solve analogical reasoning problems.^{32,33} Thus, there appears some precedence in the notion that animals might be capable of analogical

reasoning; however, both these experiments required animals to make comparisons along a single stimulus dimension: size. Questions of whether size represents a unique stimulus dimension for reasoning by analogy, or whether animals might be able to flexibly reason analogically using a variety of stimulus dimensions, are not answerable using the above experimental paradigm. Therefore, to test for analogical reasoning abilities without requiring specific token or symbol training, researchers have often employed the relational matching-to-sample (RMTS) task, as it has no prerequisites for linguistic ability. Commonly, this task involves presenting subjects with a sample pair of stimuli (although sometimes larger arrays³⁴) which are the same or different from one another, alongside two choice pairs of stimuli that either match or mismatch the relationship presented in the sample stimuli. Subjects must decide which choice pair of stimuli match the sample pair's relation. Using the RMTS task, researchers have shown analogical reasoning abilities in chimpanzees (but not rhesus monkeys),³⁵ baboons,^{34,36} and recently in two avian species: crows³⁷ and parrots.³⁸

During these latter two studies, avian subjects were presented with a target card that showed two sample stimuli. These stimuli could be the same or different according to three categories (size, shape, or color). The birds needed to select one (of two) comparison pairs (displayed on separate cards) that matched the relation present in the sample pair. During 75% of the trials, the correct pairing was exactly the same as the sample pair (Identity trials), whereas on the remaining 25% of trials, neither of the choice pairings matched identically the sample pair. Instead, the avian subject needed to select the pair that contained the same within-pair relationship as the sample pair (Relational trials). Both crows and parrots were able to spontaneously choose the correct stimuli during relational trials after having only been trained with identity trials, suggesting that they were able to use their abstract concept of "sameness" to complete the analogical reasoning task.

It is possible, however, that the birds in those two studies were using simpler mechanisms to solve the presented analogies, revealing potential problems with many comparative tests of analogical reasoning. For instance, Vonk³⁹ argued that the spatial configurations of the stimulus pairings on each card were not sufficiently varied to rule out whether the birds' success could have been due to the application of simpler perceptual, rather than relational, rules. For example, during a "color" trial in which the bird was presented with a sample card displaying two stimuli of the same color, the bird could respond using the perceptual rule of "one color in sample—choose one color in comparison"³⁹ (p. R70). Using a perceptual rule would allow the subject to avoid encoding the individual shapes on the card as two independent stimuli because the subject could more simply encode the entire card as the singular element with which to make comparisons (i.e., the subject could make first-order relational comparisons, rather than the higher-order analogy the researchers were intending to study). Indeed, entropy cues provided by variability within a stimulus arrangement have been shown to assist in both first-²² and second-order relational reasoning.⁴⁰ This has led some researchers to question the applicability of purely RMTS tasks to the question of whether animals can reason by analogy.⁴¹ Specifically, Pepperberg⁴¹ argues for an alternate approach that requires animals to make *simultaneous* evaluations of the similarities and differences of stimuli to ensure that no simpler perceptual mechanisms are governing the animals' behavior. Still, the question of which species are capable of performing analogical reasoning remains important. In their seminal paper positing relational reasoning as the point of cognitive discontinuity between humans and animals, Penn et al.¹⁰ provide examples of experiments that would be capable of demonstrating that animals are capable of performing analogical reasoning. Our study presents a test analogous to Penn et al.'s proposed Experiment 1, and provides a comparative test capable of evaluating analogical reasoning ability across animal species in a manner that requires subjects to attend to individual stimuli and the relationship between them.

We designed the first non-verbal test of analogical reasoning capable of evaluating analogical reasoning in animals. To assess the suitability and utility of this test, we first evaluated how adult humans, who we know are capable of reasoning by analogy, perform on this task (see Table 1). Adult participants were presented with abstract pictorial stimuli, on a computer screen, which varied according to an A:B::C:D relationship. On any particular trial, the participants were presented with two exemplar stimuli (sequentially) that depicted a relationship (e.g., white heart and white triangle) followed by a third probe stimulus (e.g., a white plus), and the participant was required to pick the single stimulus (e.g., a white circle) that completed the analogy (in this example, a different shape relationship; see Figure 1A) from one or more distractors. Importantly, none of the stimulus dimensions in the exemplar stimuli relevant to the comparison were present in the probe or choice stimuli. Thus, the participant could not use any identity information, but could only respond accurately on any given trial by attending to the *relational* information provided by the exemplar stimuli

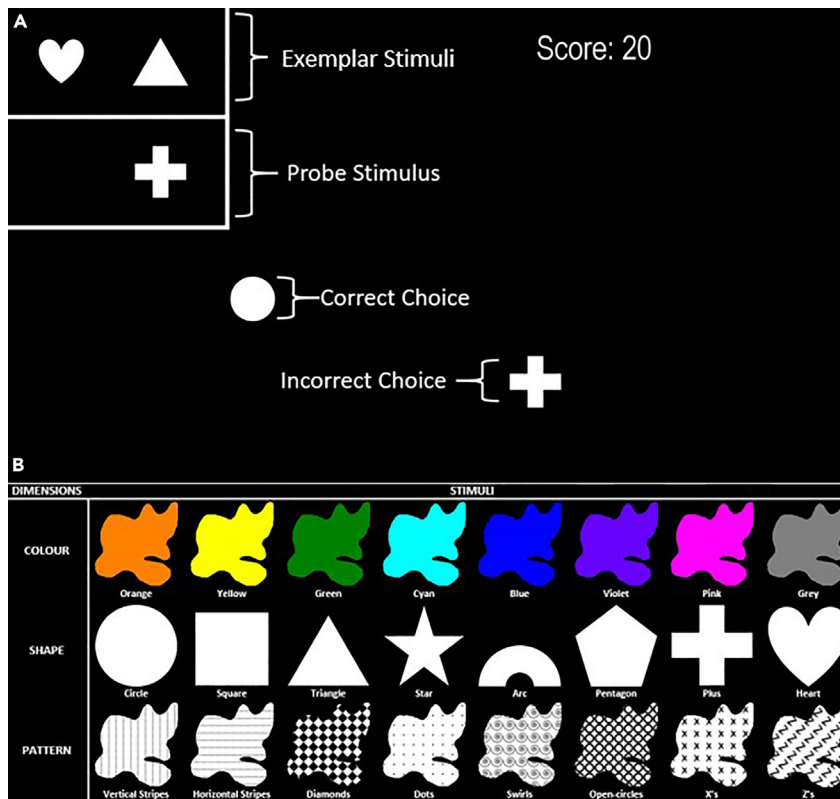


Figure 1. Experimental stimuli and procedure

(A) Example of the experiment display during First Dimension Training. Text labels, with the exception of the word “Score”, have been added for clarity and did not appear during the actual experiment. This is an example of a “Different Trial” for a participant first trained with the Shape dimension. The circle shape is correct, as it has a different shape than the probe stimulus. The “Score” represents the running total of the participant’s accumulated accuracy over the experiment.

(B) Examples of the possible colors, shapes, and patterns of stimuli. There were 8 levels of each stimuli, in addition to one “neutral” level used in earlier training phases (Color: “White”; Shape: “Amorphous”; Pattern: “Filled”).

(Table S1). To evaluate the effect of relational complexity, stimuli at the beginning of the experiment varied only along a single dimension (shape, color, or pattern) but increased in complexity (i.e., More dimensions were added that could differ between exemplars. Following the above example, the exemplar stimuli might be a white heart and a *blue* triangle). By the final phase of the experiment, the stimuli could vary along all three dimensions. During these latter trials, during which participants needed to attend to multiple stimulus dimensions, trials could be defined as more or less complex based on the number of dimensions which differed between the provided example pair (i.e., A:B), and therefore needed to be considered in order to select the correct choice stimulus. If complexity influences analogical reasoning ability, we expected participants to perform less accurately during trials with more dimensions that differed between the exemplar stimuli. To validate this procedure for comparative studies of analogical reasoning ability, participants were tested with minimal verbal instruction. They were told “Please begin. You may use the mouse to click on the objects.” In addition, participants were instructed to respond as accurately as possible to gain maximum points (see Appendix). Performance on the analogical reasoning task was compared with performance on the Raven’s Progressive Matrices, a test of fluid intelligence, to examine the relationship between these cognitive functions for humans.

RESULTS

Humans learn pictorial analogies

Participants were able to learn the task and required an average of 20.2 ± 0.86 ($M \pm SE$) trials to reach the criterion of 5 consecutive correct trials to advance to transfer testing. When required to transfer their

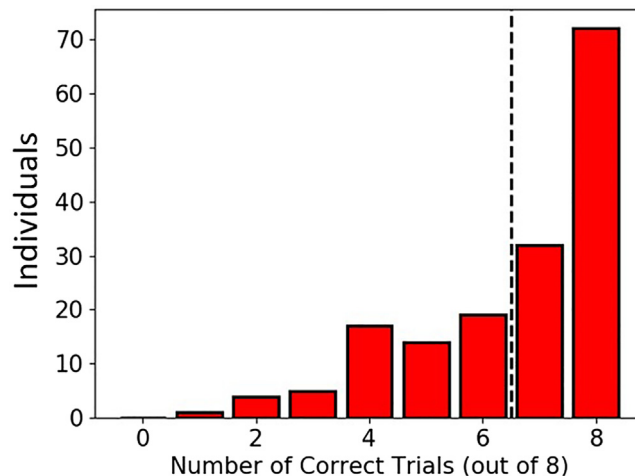


Figure 2. Distribution of the number of correct responses by number of individuals

The dashed vertical line indicates chance, as determined by a binomial response.

knowledge from one dimension onto a second, novel dimension, participants were able to do so with high accuracy (Proportion of correct transfer testing trials: 0.82 ± 0.02), indicating that they were able to reason by analogy independent of the stimulus type. However, analysis of individual participants' performance showed that the apparent overall high accuracy was due to many participants performing with ceiling-level accuracy. With eight trials, the chance probability of 0.05 for a binary choice would occur between 6 ($p = 0.109$) and 7 ($p = 0.031$) correct trials. Therefore, if we consider only scores of 7 or greater as indicative of full transfer, then only 104 (of 180) participants successfully transferred to a novel second dimension (Figure 2). At the group level, there were no significant differences in the proportion of correct transfer trials based on gender (Female = 0.81 ± 0.02 , Male = 0.84 ± 0.02), $F(1,178) = 0.769$, $p = 0.382$; nor the first dimension learned (Color = 0.81 ± 0.03 , Shape = 0.84 ± 0.03 , Pattern = 0.81 ± 0.03), $F(2,177) = 0.239$, $p = 0.787$; the interaction of gender and dimension was also not significant $F(2,177) = 0.484$, $p = 0.617$. Throughout the course of the experiment, we observed no differences based on participant gender, stimulus dimension, or the order of stimuli learned.

When the number of available dimensions increased to two, meaning that now the stimuli could differ according to two stimulus dimensions, participants were on average slightly less accurate than with a single dimension (0.74 ± 0.02). When a third dimension was introduced, but still only two dimensions could differ on any particular trial, participants maintained a high level of overall accuracy ($M = 0.80 \pm 0.01$), and finally, when all three dimensions could vary freely, participants were still able to complete the task with a high level of overall accuracy (0.82 ± 0.01). Thus, despite minimal verbal instruction, participants were able to use analogical reasoning to successfully complete all trials, up to at least a maximum of three varying stimulus dimensions.

Relational complexity reduces analogical reasoning accuracy

The "difficulty" of any given trial could be determined by the complexity of the relationship between the stimuli from which the participants needed to build their analogy. Within-relationship complexity (WRC) was defined as the number of dimensions that differed between the exemplar stimuli. When the stimuli could only differ along a single dimension, there was a significant effect of WRC, $F(1,177) = 4.13$, $p = 0.044$ (0 Dimensions Different = 0.85 ± 0.02 , 1 Dimension Different = 0.80 ± 0.02). Thus, already at this early stage in testing, participants had greater difficulty on "different trials" than "same trials". In addition, there was an effect of WRC on reaction time, $F(1,178) = 92.0$, $p < 0.001$, with participants responding faster on trials with 0 dimensions different ($M = 1.34 \pm 0.05$ s) than trials with 1 dimension different ($M = 1.94 \pm 0.08$).

When the exemplar pair were composed of two dimensions, and as a result there were three possible levels of complexity, the significant effect of WRC was even more pronounced (0 Dimensions Different = 0.90 ± 0.03 , 1 Dimension Different = 0.70 ± 0.03 , 2 Dimensions Different = 0.69 ± 0.03 ; $F(2,177) = 31.1$, $p < 0.001$). Reaction time showed a significant effect of WRC, with participants requiring more time to make a decision

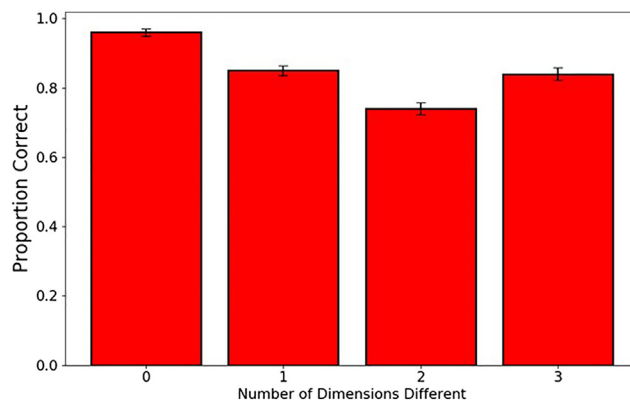


Figure 3. Proportion of correct trials during the Triple Combination Testing

The proportion correct is based on the number of dimensions differing (i.e., WRC) between the exemplar stimuli. Error bars show Standard Error of the Mean.

with additional complexity (0 Dimensions Different = 1.94 ± 0.15 s; 1 Dimension Different = 3.33 ± 0.15 s; 2 Dimensions Different = 4.22 ± 0.24 s), $F(2,177) = 69.9$, $p < 0.001$. Similarly, when the third dimension was incorporated but the exemplar stimuli still only varied according to two dimensions, there was still a significant effect of WRC on accuracy (0 Dimensions Different = 0.94 ± 0.02 , 1 Dimension Different = 0.75 ± 0.02 , 2 Dimensions Different = 0.73 ± 0.02), $F(2,177) = 77.3$, $p < 0.001$, and reaction time (0 Dimensions Different = 2.10 ± 0.09 s; 1 Dimension Different = 3.83 ± 0.15 s; 2 Dimensions Different = 4.84 ± 0.22 s), $F(2,177) = 187$, $p < 0.001$.

Lastly, when all three dimensions could vary freely, there was again a significant effect of WRC ($F(3, 176) = 58.3$, $p < 0.001$). However, unlike earlier test phases, accuracy did not decrease linearly with increased relational complexity (Figure 3). Participants showed the best accuracy when WRC was zero ($M = 0.95 \pm 0.01$), and equivalent accuracy during trials in which the exemplar stimuli had 1 dimension different ($M = 0.85 \pm 0.01$) or all 3 dimensions different ($M = 0.84 \pm 0.02$), but performed least accurately when 2 dimensions differed ($M = 0.74 \pm 0.02$). Interestingly, analysis of reaction times revealed a linear effect of relational complexity, $F(3,176) = 183$, $p < 0.001$, wherein participants took the longest to respond to trials with the greatest relational complexity (M s = 2.69, 4.74, 5.94, 6.23 s for trials with 0-, 1-, 2-, and 3-dimensions different, respectively).

Analogical performance is correlated with fluid intelligence

Overall, participants were very accurate in responding to the Raven's Progressive Matrices ($M = 0.86 \pm 0.005$). All participants completed all aspects of the experiment within the requisite time frame. However, closer examination of responses revealed that participants scored near ceiling for Sets A ($M = 0.99 \pm 0.002$) and B ($M = 0.96 \pm 0.004$), whereas there was much lower accuracy, and more variability, for responses on Set C ($M = 0.626 \pm 0.015$). Indeed, responses on several of these most difficult questions fell to chance levels (Figure 4). Because comparisons using Sets A and B would have been obfuscated by the observed ceiling effects, we decided to use only Set C scores for comparisons with performance during the computer task, as it appeared most diagnostic of problem-solving ability.

There was a significant, albeit weak, correlation between performance on Set C of the Raven's matrices, and accuracy during the first transfer test, $r = 0.234$, $p = 0.003$. Similarly, there was a significant correlation between performance on Set C and accuracy during the final, most complex phase of the experiment, wherein all three stimulus dimensions could differ, $r = 0.355$, $p < 0.001$.

DISCUSSION

During the present study, human participants were able to reason by analogy when presented with a non-verbal task, and variations in the complexity of the presented relationships influenced the difficulty in making those analogical assessments. Furthermore, we observed that analogical reasoning ability correlated with participants' fluid intelligence. These results validate that our paradigm can be used to test analogical reasoning without requiring language or real-world knowledge of concepts, making it particularly suitable

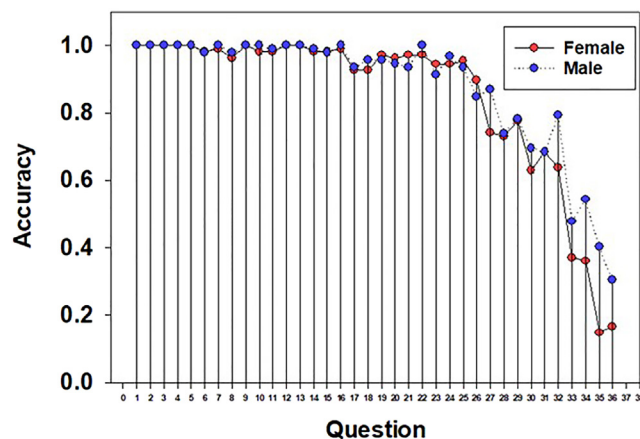


Figure 4. Proportion of correct responses for individual questions in the Raven's Progressive Matrices, Sets A (1–12), B (13–24), and C (25–36)

for use in cross-cultural, developmental, or species comparisons of analogical reasoning. Our results show that human participants were able to spontaneously complete analogies using visual stimuli that could vary with up to three dimensions. We found that accuracy during the analogy task was strongly influenced by the complexity of the relationships being compared. Accuracy decreased for trials containing relational pairs which were more complex, in that they were composed of a larger number of different relations within the pair. This result was not linear, as accuracy during the latter tests with three dimensions was higher for trials in which all three dimensions differed between the exemplar stimuli, when compared to trials in which only two dimensions differed. In addition, our results show variability in the success with which participants identified and completed the analogical rule, resulting in high levels of individual variability. Indeed, performance on the analogical reasoning task was correlated with performance on the Raven's Progressive Matrices. Taken together, these results indicate that the complexity of the relationship being analyzed has a stronger influence on the apparent difficulty of a test of analogical reasoning than does the physical (perceptual) features, or the familiarity with specific dimensions.

Participants, across all phases of the experiment, performed most accurately when the two sample stimuli were the same across all relevant dimensions, as these "same" trials simplified the task greatly; the participant needed only to select whichever choice stimulus was an exact match of the probe stimulus. There is an existing body of evidence indicating that complexity influences analogical reasoning ability. One popular test that has many similarities to the present study is the *People Pieces Analogy* (PPA) task, which previous researchers have used to demonstrate the effect of complexity on analogical reasoning ability.⁴² Participants in that study were cued to attend to a certain number of stimulus dimensions and the more dimensions to which they needed to attend, the worse their performance. This effect was most pronounced in middle- and older-aged individuals, leading the authors to suggest that the decline in reasoning performance may be explained by a decline in attention and inhibitory functions associated with age.

Unlike the PPA task, however, trials in the present study were not preceded with a description of the relationship between stimuli, meaning that the participant was required to generate the sample relationship on their own, before beginning to evaluate candidate choice stimuli. Thus, in addition to the demands of attentional and inhibitory functions, participants in the present study likely experienced a demand of working memory processes. This may explain the significant decrease in accuracy as a function of complexity, as previous research has shown that analogical reasoning requires the use of executive processes in working memory, and the ability deteriorates if interfered with by a concurrent working memory task.^{43,44} As the number of relevant stimulus dimensions increased, the status of each dimension's relationship between the sample pair would have to be held in memory during search for the choice stimulus (e.g., SAME color, DIFFERENT pattern, SAME shape). Indeed, some participants could be heard verbalizing these observations during the experiment, perhaps as an attempt to deal with increased cognitive load (personal observation). However, the present study was not designed to disentangle the competing influences of inhibition, attention, and working memory, and so future research should explore the relationship of these processes to analogical reasoning. Similarly, the present study linearly increased the potential for relational

Table 1. All levels of within-relationship complexity (WRC) for trials from Triple Combination Testing

Color	Shape	Pattern
Same	Same	Same
Same	Same	Different
Same	Different	Same
Same	Different	Different
Different	Same	Same
Different	Same	Different
Different	Different	Same
Different	Different	Different

complexity over the course of the experiment (i.e., started with only one stimulus dimension before advancing to two and finally to three). Whether participants would have responded differently had they started the experiment with the higher dimensional stimuli is not known. However, we speculate that a similar same pattern of results would be likely, as participants would still need to contend with the perceptual differences afforded by differences in relational complexity, regardless of how many stimulus dimensions are present.

Our finding that relational reasoning ability correlated with performance on the most difficult problems (Set C) of the Raven's matrices (RPM) provides further support to the idea that increased relational complexity results in an increase in cognitive load. Performance on the RPM is generally argued to be related to fluid intelligence and is dependent in part on working memory capacity. More difficult questions within the RPM can be characterized by having more abstract rules than less difficult questions, but also by the increased number of incremental steps involved in solving them.²⁰ Carpenter et al.²⁰ argued that solving the more complex RPM problems requires breaking down the larger visual problem into smaller subcomponents (what Carpenter refers to as "tokens") which are simpler to process, and incrementally addressing each token. They argued that the more challenging RPM questions were composed of a higher number of subcomponents that needed to be systematically processed, in turn, suggesting that the limiting factor in this chain is working memory capacity. This process of decomposition-processing-recomposition is also necessary to solve our analogical task, particularly in the later phases of testing where the number of relevant stimuli dimensions increase. Alternatively, subsequent experiments which controlled for working memory capacity found that RPM performance is not dependent strictly on working memory capacity, but rather on the ability of the individual to control their attention and eliminate items of distraction and interference.^{45,46} Although we did not test specifically whether working memory capacity or attentional components might have led to greater errors, all of the relevant stimuli (i.e., the two exemplars and the probe) for any particular trial were available on the screen when the participant needed to make a choice. This along with the fact that participants were at most required to keep three items (shape, color, and pattern) in working memory suggests that working memory capacity was not likely the main source of errors in the present study. Instead, the ability to adequately attend to the relevant stimuli dimensions was most important in accurate responding, as shown by high overall accuracy, even with greater numbers of stimulus dimensions. Using a novel set of relational matrices, Duncan et al.⁴⁷ identified attentional control functions as key in these kinds of tests of fluid intelligence. When solving complex problems commonly found in fluid intelligence tests, participants need to deconstruct the task into simpler components, to which they must individually attend. Attention, then, plays a key part in performance during the present computer-based experiment, as the stimuli used in this experiment can similarly be deconstructed into individual (in our case, dimensional) components.

Multiple different brain regions have been implicated in tests of fluid intelligence and analogical reasoning. The multiple-demand system, that includes, among other regions, the lateral frontal surface, anterior insula, and dorsomedial frontal cortex, has been found to be involved in tests of cognitive control and fluid intelligence due to those regions' importance in attentional control.²¹ The importance of attention and abstraction has long led researchers to implicate the prefrontal cortex (PFC) in analogical reasoning ability.⁴⁸ Specifically, Holyoak and Kroger⁴⁸ argued that the PFC evolved to deal with the analysis of "relationships at increasing levels of complexity" (p. 257). These theoretical considerations have subsequently been validated with experimental observations, as individuals with damage to their prefrontal lobes demonstrate selective deficits to their relational reasoning ability.⁴⁹ Subjects with prefrontal lobe damage

exhibited no decrements to episodic memory or semantic knowledge, whereas subjects with intact prefrontal cortex (but damage to anterior temporal regions) showed no deficits to their analogical reasoning ability. This suggests the PFC is essential for the type of relational processing required by analogies. Results from brain imaging experiments further clarify this, with participants performing analogy problems under fMRI demonstrating greater activation in the PFC, particularly the left anterior and rostral PFC.^{50,51} Children show an ability to perform relational reasoning as young as 4 and 5 years old⁵² and the development of their analogical ability is correlated with increased engagement of the left anterior PFC.⁵³ This has led some researchers to suggest that one major point of discontinuity between human and animal cognition is related to differences in prefrontal cortices.¹⁰ These neurological differences, as well as differences in language capacity, have long led analogy to be considered capable by only humans, or language- or token-trained non-human primates.⁵⁴ However, recent evidence has shown that analogical problem solving by monkeys (capuchins;³³ baboons^{34,36}) is not limited by lexical training nor symbolic representation. Similarly, the ability to recognize relational similarity by apes is not dependent on linguistic training,³² suggesting that analogical reasoning abilities may be more widely distributed across species than previously thought.

Recent work with avian subjects has broadened the conversation on which species might be capable of analogical reasoning even further. Birds, although lacking a PFC, possess a corresponding structure—the nidopallium caudolaterale—that has anatomical and functional similarities to the PFC,⁵⁵ including governing tasks requiring executive control and in particular working memory.⁵⁶ Using a relational matching-to-sample paradigm, crows³⁷ and parrots³⁸ showed an ability to perform analogical reasoning. Subsequent commentary, however, has questioned the strength of these initial findings.^{39,57} Specifically, it is possible that rather than using the intended relational rules, subjects could have used simpler perceptual or configural cues to *appear* to be responding relationally.³⁹ Such alternative strategies would not be possible in the present study, as participants needed to find and select the single stimulus that would complete the pattern from stimuli that were not physically adjacent to the probe stimulus (as was the case in the study by Smirnova et al.³⁷). Thus, participants could not use within-pair variability or entropy to respond accurately, and therefore had to use a relational rule. In addition, the relational reasoning task used to test parrots and crows was simpler than the one used in the present study, in that it employed a binary choice, and therefore could not measure the limits of a subject's analogical reasoning ability; subjects either could or could not reason analogically. Using this approach does not permit one to determine at which levels of complexity an individual might be able to perform analogical reasoning, and at which levels they might not. The present study has the advantage of doing so in that even incorrect responding can elucidate precisely which cues subjects are using, when not using relational cues. That is, our method allows us to detect if participants respond on the basis of familiarity (that is, showing biases toward previously learned stimulus dimensions) or perceptual similarity by examining their incorrect choices and evaluating which (if any) of the individual dimensions were correct. For example, subjects might show higher accuracy for first stimulus dimensions learned, which would indicate limitations in the flexibility of their analogical ability. Alternatively, they could show higher accuracies for one particular stimulus dimension over another, which might indicate natural perceptual biases, perhaps even on a species level. Overall, however, subjects in the present study must not simply *recognize* analogical relationships, as in previous studies, but must *construct* and *complete* them independently. This is a more stringent requirement than that used in many previous studies, but may more unambiguously determine which species are capable of using relational reasoning.

Species within the families Corvidae and Psittacidae remain some of the best candidates for investigating the ability of animals to understand analogies, given their evidence of complex cognition in a variety of relevant areas. However, our paradigm can be valuable for evaluating the relational reasoning abilities of a range of species using a more stringent test, wherein configural strategies cannot lead to accurate responding (as criticized by Vonk³⁹ for prior tests of analogical reasoning by animals). Future research employing our procedure could determine whether factors of complexity or familiarity influence the rudiments of analogical reasoning across a range of species. Because the proposed method relies only on the ability to interact with a touchscreen monitor, it may be applied not only to the parrots, corvids, and primates previously studied but also a diverse breadth of other species.

In conclusion, the present study developed and tested a non-verbal task for analyzing analogical reasoning. The task was designed to allow direct cross-species comparisons on analogical reasoning ability. Here, we tested and validated the procedure using adult human participants. Participants given

minimal verbal instructions were able to solve the analogical reasoning task. Analogical reasoning ability was found to be correlated with fluid intelligence, as measured by the Raven's progressive matrices, and relational complexity was found to influence analogical ability. Thus, the methods here produce similar patterns of results reported in previous research on human analogical reasoning ability, indicating construct validity. This provides a benchmark of human performance, against which the performance of animals may be compared, broadening our understanding of the importance and prevalence of relational reasoning.

Limitations of the study

The main limitation of the present study is that, although it provides a methodology for comparative studies of analogy across species, we have only evaluated the procedure with human participants. Modifications to the experimental procedures outlined here might be necessary to conduct the experiment with non-human animals (e.g., to accommodate species differences in color perception). In addition, because relational complexity increased linearly over the course of the experiment, we are unable to assess how participants may have responded if given complex relationships at the onset of the experiment.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Participants
- METHOD DETAILS
 - Experimental Stimuli
 - General procedures
 - Analogical reasoning task
 - Ishihara colour-blindness test
 - Raven's 2 progressive matrices
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.106392>.

ACKNOWLEDGMENTS

We thank Iroshini Gunasekera and Nicole Tongol for their assistance in data collection, and Randy Jamieson for his invaluable suggestions and insights. This research was supported by a Natural Science and Engineering Research Council of Canada (NSERC) Discovery grant (#RGPIN/4944-2017) and an NSERC Canada Research Chair to DMK.

AUTHOR CONTRIBUTIONS

Conceptualization, K.L., D.K.; Methodology, K.L., D.K.; Software, K.L.; Data Collection, K.L.; Formal Analysis, K.L.; Writing – Original Draft, K.L., D.K.; Writing – Review & Editing, K.L., D.K.; Visualization, K.L.; Supervision, D.K.; Project Administration, D.K.; Funding Acquisition, D.K.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location. We support inclusive, diverse, and equitable conduct of research.

Received: May 12, 2022
Revised: August 17, 2022
Accepted: March 8, 2023
Published: March 11, 2023

REFERENCES

- Bod, R. (2009). From exemplar to grammar: a probabilistic analogy-based model of language learning. *Cogn. Sci.* 33, 752–793. <https://doi.org/10.1111/j.1551-6709.2009.01031.x>.
- Brand, C.O., Mesoudi, A., and Smaldino, P.E. (2021). Analogy as a catalyst for cumulative cultural evolution. *Trends Cogn. Sci.* 25, 450–461. <https://doi.org/10.1016/j.tics.2021.03.002>.
- Zemeckis, R., Director (1994). *Forrest Gump* [Film] (The Tisch Company).
- Gick, M.L., and Holyoak, K.J. (1983). Schema induction and analogical transfer. *Cogn. Psychol.* 15, 1–38. [https://doi.org/10.1016/0010-0285\(83\)90002-6](https://doi.org/10.1016/0010-0285(83)90002-6).
- Burt, C. (1949). The structure of the mind: a review of the results of factor analysis. *Br. J. Educ. Psychol.* 19, 176–199. <https://doi.org/10.1111/j.2044-8279.1949.tb01621.x>.
- Hofstadter, D.R. (2001). Analogy as the core of cognition. In *The analogical mind: perspectives from cognitive science*, pp. 499–538.
- Christie, S., and Gentner, D. (2014). Language helps children succeed on a classic analogy task. *Cogn. Sci.* 38, 383–397. <https://doi.org/10.1111/cogs.12099>.
- Spellman, B.A., and Holyoak, K.J. (1992). If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles. *J. Pers. Soc. Psychol.* 62, 913–933. <https://doi.org/10.1037/0022-3514.62.6.913>.
- Goswami, U., and Brown, A.L. (1990). Melting chocolate and melting snowmen: analogical reasoning and causal relations. *Cognition* 35, 69–95. [https://doi.org/10.1016/0010-0277\(90\)90037-K](https://doi.org/10.1016/0010-0277(90)90037-K).
- Penn, D.C., Holyoak, K.J., and Povinelli, D.J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* 31, 109–130. , discussion 130–178. <https://doi.org/10.1017/S0140525X08003543>.
- Ichien, N., Lu, H., and Holyoak, K.J. (2020). Verbal analogy problem sets: an inventory of testing materials. *Behav. Res. Methods* 52, 1803–1816. <https://doi.org/10.3758/s13428-019-01312-3>.
- Gick, M.L., and Holyoak, K.J. (1980). Analogical problem solving. *Cogn. Psychol.* 12, 306–355. [https://doi.org/10.1016/0010-0285\(80\)90013-4](https://doi.org/10.1016/0010-0285(80)90013-4).
- Sternberg, R.J. (1977). Component processes in analogical reasoning. *Psychol. Rev.* 84, 353–378. <https://doi.org/10.1037/0033-295X.84.4.353>.
- Duncker, K. (1945). On problem solving. *Psychol. Monogr.* 58. <https://doi.org/10.1037/h0093599>.
- Chan, J., Paletz, S.B.F., and Schunn, C.D. (2012). Analogy as a strategy for supporting complex problem solving under uncertainty. *Mem. Cognit.* 40, 1352–1365. <https://doi.org/10.3758/s13421-012-0227-z>.
- Goswami, U. (1989). Relational complexity and the development of analogical reasoning. *Cogn. Dev.* 4, 251–268. [https://doi.org/10.1016/0885-2014\(89\)90008-7](https://doi.org/10.1016/0885-2014(89)90008-7).
- Smith, J.D., Jackson, B.N., and Church, B.A. (2019). Breaking the perceptual-conceptual barrier: relational matching and working memory. *Mem. Cognit.* 47, 544–560. <https://doi.org/10.3758/s13421-018-0890-9>.
- Lovett, A., Forbus, K., and Usher, J. (2007). Analogy with qualitative spatial representations can simulate solving Raven's Progressive Matrices. *Proc. Annu. Meet. Cognit. Soc.* 29.
- Little, D.R., Lewandowsky, S., and Craig, S. (2014). Working memory capacity and fluid abilities: the more difficult the item, the more more is better. *Front. Psychol.* 5, 239. <https://doi.org/10.3389/fpsyg.2014.00239>.
- Carpenter, P.A., Just, M.A., and Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven progressive matrices test. *Psychol. Rev.* 97, 404–431. <https://doi.org/10.1037/0033-295X.97.3.404>.
- Duncan, J., Assem, M., and Shashidhara, S. (2020). Integrated intelligence from distributed brain activity. *Trends Cogn. Sci.* 24, 838–852.
- Wasserman, E.A., Fagot, J., and Young, M.E. (2001). Same-different conceptualization by baboons (*Papio papio*): the role of entropy. *J. Comp. Psychol.* 115, 42–52. <https://doi.org/10.1037/0735-7036.115.1.42>.
- Wasserman, E.A., and Young, M.E. (2010). Same-different discrimination: the keel and backbone of thought and reasoning. *J. Exp. Psychol. Anim. Behav. Process.* 36, 3–22. <https://doi.org/10.1037/a0016327>.
- Katz, J.S., Wright, A.A., and Bachevalier, J. (2002). Mechanisms of same/different abstract-concept learning by rhesus monkeys (*Macaca mulatta*). *J. Exp. Psychol. Anim. Behav. Process.* 28, 358–368. <https://doi.org/10.1037/0097-7403.28.4.358>.
- Magnotti, J.F., Katz, J.S., Wright, A.A., and Kelly, D.M. (2015). Superior abstract-concept learning by Clark's nutcrackers (*Nucifraga columbiana*). *Biol. Lett.* 11, 20150148. <https://doi.org/10.1098/rsbl.2015.0148>.
- Pepperberg, I.M. (1987). Acquisition of the same different concept by an African gray parrot (*Psittacus-Erithacus*) - learning with respect to categories of colour, shape, and material. *Anim. Learn. Behav.* 15, 423–432. <https://doi.org/10.3758/BF03205051>.
- Katz, J.S., and Wright, A.A. (2006). Same/Different abstract-concept learning by pigeons. *J. Exp. Psychol. Anim. Behav. Process.* 32, 80–86. <https://doi.org/10.1037/0097-7403.32.1.80>.
- Wasserman, E.A., Castro, L., and Freeman, J.H. (2012). Same-different categorization in rats. *Learn. Mem.* 19, 142–145. <https://doi.org/10.1101/lm.025437.111>.
- Avarguès-Weber, A., Dyer, A.G., Combe, M., and Giurfa, M. (2012). Simultaneous mastering of two abstract concepts by the miniature brains of bees. *Proc. Natl. Acad. Sci. USA* 109, 7481–7486. <https://doi.org/10.1073/pnas.1202576109>.
- Premack, D. (1976). Intelligence in Ape and Man (Erlbaum).
- Oden, D.L., Thompson, R.K., and Premack, D. (2001). Can an ape reason analogically? Comprehension and production of analogical problems by Sarah, a chimpanzee (*Pan troglodytes*). In *The analogical mind: Perspectives from cognitive science*, D. Gentner, K.J. Holyoak, and B.N. Kokinov, eds. (The MIT Press), pp. 471–497.
- Haun, D.B.M., and Call, J. (2009). Great apes' capacities to recognize relational similarity. *Cognition* 110, 147–159. <https://doi.org/10.1016/j.cognition.2008.10.012>.
- Kennedy, E.H., and Frigaszy, D.M. (2008). Analogical reasoning in a capuchin monkey (*Cebus apella*). *J. Comp. Psychol.* 122, 167–175. <https://doi.org/10.1037/0735-7036.122.2.167>.
- Flemming, T.M., Thompson, R.K.R., and Fagot, J. (2013). Baboons, like humans, solve analogy by categorical abstraction of relations. *Anim. Cogn.* 16, 519–524. <https://doi.org/10.1007/s10071-013-0596-0>.
- Flemming, T.M., Beran, M.J., Thompson, R.K.R., Kleider, H.M., and Washburn, D.A. (2008). What meaning means for same and different: analogical reasoning in humans (*Homo sapiens*), chimpanzees (*Pan troglodytes*), and rhesus monkeys (*Macaca mulatta*). *J. Comp. Psychol.* 122, 176–185. <https://doi.org/10.1037/0735-7036.122.2.176>.
- Fagot, J., and Parron, C. (2010). Relational matching in baboons (*Papio papio*) with reduced grouping requirements. *J. Exp. Psychol. Anim. Behav. Process.* 36, 184–193. <https://doi.org/10.1037/a0017169>.

37. Smirnova, A., Zorina, Z., Obozova, T., and Wasserman, E. (2015). Crows spontaneously exhibit analogical reasoning. *Curr. Biol.* 25, 256–260. <https://doi.org/10.1016/j.cub.2014.11.063>.
38. Obozova, T., Smirnova, A., Zorina, Z., and Wasserman, E. (2015). Analogical reasoning in amazons. *Anim. Cogn.* 18, 1363–1371. <https://doi.org/10.1007/s10071-015-0882-0>.
39. Vonk, J. (2015). Corvid cognition: something to crow about? *Curr. Biol.* 25, R69–R71. <https://doi.org/10.1016/j.cub.2014.12.001>.
40. Fagot, J., Wasserman, E.A., and Young, M.E. (2001). Discriminating the relation between relations: the role of entropy in abstract conceptualization by baboons (*Papio papio*) and humans (*Homo sapiens*). *J. Exp. Psychol. Anim. Behav. Process.* 27, 316–328.
41. Pepperberg, I.M. (2021). How do a pink plastic flamingo and a pink plastic elephant differ? Evidence for abstract representations of the relations same-different in a Grey parrot. *Curr. Opin. Behav. Sci.* 37, 146–152. <https://doi.org/10.1016/j.cobeha.2020.12.010>.
42. Viskontas, I.V., Morrison, R.G., Holyoak, K.J., Hummel, J.E., and Knowlton, B.J. (2004). Relational integration, inhibition, and analogical reasoning in older adults. *Psychol. Aging* 19, 581–591. <https://doi.org/10.1037/0882-7974.19.4.581>.
43. Morrison, R., Holyoak, K., and Truong, B. (2001). Working-memory modularity in analogical reasoning. In *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (Erlbaum), pp. 663–668.
44. Waltz, J.A., Lau, A., Grewal, S.K., and Holyoak, K.J. (2000). The role of working memory in analogical mapping. *Mem. Cognit.* 28, 1205–1212. <https://doi.org/10.3758/BF03211821>.
45. Guarino, K.F., Wakefield, E.M., Morrison, R.G., and Richland, L.E. (2021). Exploring how visual attention, inhibitory control, and co-speech gesture instruction contribute to children's analogical reasoning ability. *Cognit. Dev.* 58, 101040. <https://doi.org/10.1016/j.cogdev.2021.101040>.
46. Unsworth, N., and Engle, R. (2005). Working memory capacity and fluid abilities: examining the correlation between Operation Span and Raven. *Intelligence* 33, 67–81.
47. Duncan, J., Chylinski, D., Mitchell, D.J., and Bhandari, A. (2017). Complexity and compositionality in fluid intelligence. *Proc. Natl. Acad. Sci. USA* 114, 5295–5299. <https://doi.org/10.1073/pnas.1621147114>.
48. Holyoak, K.J., and Kroger, J.K. (1995). Forms of reasoning: insight into prefrontal functions. *Ann. N. Y. Acad. Sci.* 769, 253–263. <https://doi.org/10.1111/j.1749-6632.1995.tb38143.x>.
49. Waltz, J.A., Knowlton, B.J., Holyoak, K.J., Boone, K.B., Mishkin, F.S., de Menezes Santos, M., Thomas, C.R., and Miller, B.L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychol. Sci.* 10, 119–125. <https://doi.org/10.1111/1467-9280.00118>.
50. Krawczyk, D.C., McClelland, M.M., Donovan, C.M., Tillman, G.D., and Maguire, M.J. (2010). An fMRI investigation of cognitive stages in reasoning by analogy. *Brain Res.* 1342, 63–73. <https://doi.org/10.1016/j.brainres.2010.04.039>.
51. Krawczyk, D.C. (2012). The cognition and neuroscience of relational reasoning. *Brain Res.* 1428, 13–23. <https://doi.org/10.1016/j.brainres.2010.11.080>.
52. Rattermann, M.J., and Gentner, D. (1998). The effect of language on similarity: the use of relational labels improves young children's performance in a mapping task. *Cognit. Dev.* 13, 453–478.
53. Whitaker, K.J., Vendetti, M.S., Wendelken, C., and Bunge, S.A. (2018). Neuroscientific insights into the development of analogical reasoning. *Dev. Sci.* 21, e12531. <https://doi.org/10.1111/desc.12531>.
54. Thompson, R.K., Oden, D.L., and Boysen, S.T. (1997). Language-naïve chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *J. Exp. Psychol. Anim. Behav. Process.* 23, 31–43. <https://doi.org/10.1037/0097-7403.23.1.31>.
55. Güntürkün, O. (2005). The avian 'prefrontal cortex' and cognition. *Curr. Opin. Neurobiol.* 15, 686–693. <https://doi.org/10.1016/j.conb.2005.10.003>.
56. Rose, J., and Colombo, M. (2005). Neural correlates of executive control in the avian brain. *PLoS Biol.* 3, e190. <https://doi.org/10.1371/journal.pbio.0030190>.
57. Dymond, S., and Stewart, I. (2016). Relational and analogical reasoning in comparative cognition. *Int. J. Comp. Psychol.* 29. <https://escholarship.org/uc/item/2225s37c>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Data from "Relational Complexity Influences Analogical Reasoning Ability"	This paper, deposited in Zenodo	https://doi.org/10.5281/zenodo.7212949 .
Software and algorithms		
SUMO Paint software	www.Python.org	N/A
Other		
Human Participants	This paper	Study Protocol Number: #HS22741 (University of Manitoba - Psychology/Sociology Research Ethics Board)

RESOURCE AVAILABILITY

Lead contact

Further requests for information, data, and code should be directed to the Lead Contact, Debbie M. Kelly (Debbie.kelly@umanitoba.ca).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data: The data for this study are available at: <https://doi.org/10.5281/zenodo.7212949>.
- Code: Reasonable requests for coding used in this study is available by the corresponding author upon request.
- All Other Items: Any additional information is available by the corresponding author upon request. Accession codes are not applicable for this study.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants

Participants were undergraduate Psychology students (n = 180, 90 self-identified as female and 90 self-identified as male) from the University of Manitoba. Participants ranged in age from 18 to 30 years. Students volunteered through the online participant system, SONA. Written informed consent was obtained by all participants. Students were compensated for their time with participation credits towards their final grade in the PSYC1200 course. All participants were required to have normal or corrected-to-normal vision. All procedures were carried out under the approval of the University of Manitoba Psychology/Sociology Research Ethics Board (Protocol Number: HS22741). Informed consent was obtained from all participants.

METHOD DETAILS

Participants experienced three tasks during the course of the experiment: an analogical reasoning task, a colour blindness test, and the Raven's progressive matrices. The analogical reasoning task was conducted on a desktop computer, with stimuli presented to the participant via a monitor (1920x1080, LG Flatron W2442PA). The computer program was written in Python 3. During the experiment, the overhead lights in the room were extinguished and the room was dimly lit by a desk lamp (40 watts) to allow for better salience of the visual stimuli. During the analogical reasoning task, participants sat with their chin rested on a chin rest (135cm height, 70cm from screen) to maintain a consistent viewing angle across participants. Participants interacted with the computer program via a computer mouse, situated on their preferred side.

Experimental Stimuli

All stimuli in the experiment were designed using SUMO Paint software. The stimuli were all approximately the same size (surface area ranged between 6200 – 7600 pixels). Stimuli were created to vary along three dimensions: Colour, Shape, and Pattern. For each dimension we created 8 variations (i.e. 8 different colours, 8 different shapes, and 8 different patterns), resulting in 512 possible unique stimuli. Additionally, we created a single variation as a *Neutral* exemplar of each dimension (i.e., Colour: "White"; Shape: "Amorphous"; Pattern: "Filled"). These Neutral exemplars were used to generate stimuli which corresponded to a specific dimension before the participant encountered variability within that dimension (See [Figure 1B](#) for a list of stimuli dimensions).

General procedures

Participants were allotted 65 minutes maximum to complete the entire experimental session: 30 minutes maximum for the analogical reasoning task, then approximately 5 minutes for the Ishihara colour blindness test, and a final 30 minutes maximum for the Raven's 2 Progressive Matrices (Sets A, B, C). Each participant was assigned to a different group, which determined the sequence in which they would be presented with the stimulus dimensions during the analogical reasoning task (6 groups: e.g. Colour-Shape-Pattern, Colour-Pattern-Shape, Pattern-Colour-Shape, Pattern-Shape-Colour, Shape-Colour-Pattern, Shape-Colour-Pattern). At the beginning of the session, the experimenter met with the participant, explained the general sequence of events, answered any general questions, and asked them to sign an informed consent form. The experiment began when the participant was seated in the chair and placed their chin on the chin rest.

Analogical reasoning task

Once the participant was seated in the chair, they were read a script of instructions (see [Appendix 1](#)). The instructions were designed to provide the minimum required information to the participant to more closely mirror the methodology of operant experiments with non-human animals and to avoid potentially cueing the participant that the experiment involved the topic of analogy. Importantly, the participant was instructed to respond as accurately as possible to gain maximum points (see below) during the experiment. The experimenter then initiated the program and sat behind the participant in the same experimental room, out of sight of the participant.

First Dimension Training

Trials were presented in a consistent format. At the beginning of each trial, two white rectangular *presentation* frames (10 px line width) on a black background were presented in the upper left-hand corner of the screen. One stimulus would be presented in either the left or right half (counterbalanced) of the upper white frame. The stimulus would be neutral along two dimensions and varied along the participant's training dimension (e.g. stimuli for participants starting with Shape would be white, with the filled pattern but be varying along the shape dimension). The participant was required to use the mouse to click on this stimulus, which triggered a second stimulus to appear in the other half of the upper white frame. This newly presented stimulus would either be identical to the first stimulus (Same Trial) or would be different along the training dimension (Different Trial). Together, these first two stimuli composed the *sample pair* that the participants would use as the basis for the relational assessment for that trial. Clicking on the second *sample* stimulus caused a new stimulus to appear in the lower white frame in either the left or right position (counterbalanced). This was the *probe* stimulus, which composed one half of the relationship which the participants needed to create. The probe stimulus was never identical to either of the sample stimuli, meaning that the only relevant characteristics of the sample stimuli that the participant could use was the relationship between the sample stimuli themselves. During the first four trials, clicking the *probe* stimulus caused one *choice* stimulus to appear in the larger area of the screen outside of the *presentation* frames. The *choice* stimulus would be related to the probe stimulus in the same relational manner as existed within the *sample pair* (Correct Stimulus; e.g. during a "Same trial", the correct *choice* stimulus would be the same as the probe stimulus). During all trials from the fifth trial onward, clicking the *probe* stimulus caused two *choice* stimuli to appear, the correct stimulus as described above, as well as another *choice* stimulus (Incorrect Stimulus) which would be incorrect according to the trained dimension (e.g. during a "Same trial", the incorrect stimulus would be different from the probe stimulus along the training dimension; see [Figure 1](#)). When the participant clicked either of the *choice stimuli*, the screen went black and a large message displayed the score the participant received for their choice: "+5" or "+0" for correct and

incorrect choices, respectively, for the duration of a 3 second inter-trial interval. The updated score would be reflected in the participant's cumulative total during subsequent trials in the upper right corner of the screen. If the participant chose the incorrect stimulus, they received a correction procedure in which they repeated the same trial. Once the participant had completed a minimum of 8 trials, and completed 5 consecutive trials correctly, they would advance to Transfer Testing.

First transfer testing

Eight transfer test trials were conducted to examine whether participants could transfer their learning of the analogical rule to a novel stimulus dimension. Transfer test trials were conducted in the same manner as training trials, with the exception that all of the stimuli were now varying along one of the other two dimensions and the trained dimension was neutral along all test trials (e.g. participants trained with Shape as their first dimension would be tested using either Color or Pattern; counterbalanced). Additionally, transfer test trials, and all subsequent test trials, were conducted without reinforcement. Once the participant completed all transfer test trials, they automatically advanced to second dimension training, independent of their accuracy during testing.

Second dimension training

The second dimension training trials were identical to those from First Dimension Training, but were now conducted with the new stimulus dimension which the participants encountered during the Transfer Testing. This training was conducted to ensure that participants were able to discriminate relational differences along the second dimension. As in training for the first dimension, participants advanced to the subsequent testing phase when they completed a minimum of 8 trials and completed 5 consecutive trials correctly.

Dual Dimension Transfer Testing

Eight dual dimension test trials evaluated whether participants could discriminate relationships of increased complexity. During these trials, all stimuli were constructed using the two stimulus dimensions encountered thus far. Note that this resulted in novel degrees of relationships that the *sample* pair could be related. Specifically, the two stimuli in the *sample* pair could be Same-Same, Same-Different, Different-Same, or Different-Different along the First-Second dimension, respectively. This was defined as the within-relationship complexity, as it was assumed that trials in which stimuli were the same across all dimensions would be easier to solve than those in which they were different across all dimensions. During combination testing, participants received two trials in each variation of WRC. Additionally, to account for the increased number of ways that the stimuli could be defined along 2 dimensions, the number of *choice* stimuli was increased to four: One correct along both dimensions, one correct along the first dimension but not the second, one correct along the second dimension but not the first, and one incorrect along both dimensions. This allowed us to examine whether there were particular dimensions that participants tended to use when making relational assessments, as well as how relational complexity influenced this ability. Once the participant completed all Dual Dimension Transfer Testing trials, they automatically advanced to Dual Dimension Training.

Dual Dimension Training

Dual Dimension training was conducted using the same stimuli dimensions presented in Dual Dimension testing. Additionally, because there were now four choice stimuli, the feedback of reinforcement that the participants received was modified from the First Dimension and the Second Dimension training. A fully correct response still elicited a message of "+5" (points), and a fully incorrect response "+0", but now choices which were correct along one dimension but incorrect along the other elicited a message of "+2" to indicate to the participant that they were partially correct. Anything other than a fully correct response resulted in a correction trial. Participants advanced to the subsequent testing phase when they completed a minimum of 8 trials and completed 5 consecutive trials correctly on their first attempt.

Third dimension transfer testing

This phase of testing allowed us to examine whether participants could transfer to a new stimulus dimension without increasing the complexity of the stimulus relationship. To control for the potentially confounding influence of complexity that comes from adding a third stimulus dimension, on any given test trial during this phase only two of the three stimulus dimensions varied between stimuli and were therefore relevant

in the analogical assessment; the third dimension was randomly selected and held constant across *all* stimuli during the trial (e.g. if the third dimension was shape, all stimuli would share the same shape). This maintained the relational complexity the same as during the Dual Dimension testing. Participants again received 2 trials for each level of WRC, with the addition of counterbalancing for each combination of two relevant stimulus dimensions (i.e. Colour-Shape, Colour-Pattern, Shape-Pattern), for a total of 24 test trials. Once the participant completed all test trials, they automatically advanced to training that incorporated the third dimension.

Third dimension transfer training

Training with the third dimension was conducted identically to previous training phases, with the stimuli as in Third Dimension Testing. Which two of the three stimulus dimensions were relevant for any particular trial was determined randomly. Participants advanced to the final testing phase when they completed a minimum of 8 trials and completed 5 consecutive trials correctly.

Triple combination testing

This final phase of testing allowed us to examine whether participants could make analogical assessments with an increased level of relational complexity. During this phase of testing, all three stimulus dimensions could vary, and were all relevant for responding correctly on each trial. With three free-to-vary dimensions, 8 different variations of WRC were possible based on how many (and which) dimensions were the same or different within the exemplar relationship (Table 1). Participants received 3 trials for each of the 8 variations of WRC, for a total of 24 test trials. During any particular trial, there were still four choice stimuli available: one that was correct along all three dimensions, one that was correct along two dimensions but incorrect along the third, one correct along one dimension but incorrect along the other two, and one that was incorrect along all three dimensions. The frequency with which particular stimulus dimensions (i.e. colour, shape, pattern) were correct/incorrect was counterbalanced across trials. Once the participant completed all test trials, a completion message was displayed on the screen, the analogical reasoning task was ended, and the participants proceeded to complete the Ishihara test.

Ishihara colour-blindness test

The Ishihara colour-blindness test was conducted on the same computer screen as the analogical reasoning task. The Ishihara test is used to test for deficiencies in colour-vision ability. The participant kept their chin on the chinrest, and was presented with 15 slides of numbers composed of small dots of varying colours. Participants were asked to respond quickly (within 3 seconds) with what number they saw on the slide. A minimum score of 13/15 correct was required to show sufficient colour vision for a participant to have their data included in the results. The Ishihara test was administered after the computer task so as not to cue the participants to the idea that colours might be relevant for that task. Following the Ishihara test, participants completed the final component of the experiment, the Raven's 2 progressive matrices.

Raven's 2 progressive matrices

The Raven's 2 Progressive Matrices (henceforth referred to as Raven's 2) is a standardized, non-verbal test of "deductive" reasoning and fluid intelligence. The test consists of multiple-choice items, wherein each item presents an abstract pattern composed of a matrix of cells. All cells, except for one, contain a segment of the pattern. The participant is required to identify the missing segment from the choice items that would best complement the pattern. We chose to administer the Raven's 2, in paper format, as a measure of external validation of our computer task, as it is the most popular test of general intelligence and fluid intelligence (Gignac, 2015), involving elements of visual problem solving and working memory (Carpenter, Just, & Shell, 1990). Thus, we expected that performance during the computer task would be correlated with performance on the Raven's 2. Problem sets in the Raven's 2 steadily increase in complexity, so we asked participants to complete sets A, B, and C. Each set is composed of 12 items. Participants were allowed a maximum of 30 minutes to complete the three sets of problems. Once participants finished, they were debriefed on the true purpose of the computer task as a test of analogical reasoning.

QUANTIFICATION AND STATISTICAL ANALYSIS

For First Dimension training, we conducted an ANOVA on the number of trials required to reach criterion, with Gender and First Dimension Learned as between-subjects factors, to see if either gender, or the

physical properties of certain stimulus dimensions allowed for easier relational assessments. For each testing phase, we measured the proportion of correct responses. Due to the multi-dimensional nature of the stimuli, "correct" could be identified in two ways: (1) Overall Correct – meaning the stimulus selected was the single correct stimulus available, and (2) Dimensionally Correct – meaning the stimulus selected was correct along a particular stimulus dimension (e.g. colour), but not necessarily correct along the other available dimensions. We conducted ANOVAs based on the proportion of overall correct responses at each test level (i.e., overall accuracy), gender and dimension order as within-subjects factors. Similar ANOVAs were conducted on the proportion of dimensionally correct responses (i.e., dimensional accuracy). Additionally, we examined whether WRC would explain differences in performance. We predicted that as WRC increased, as defined by a greater proportion of "Different" relationships between the *sample* pair, overall accuracy would decrease. ANOVAs were conducted on the number of relations that differed (i.e., 0, 1, 2, or 3). Additionally, ANOVAs on the reaction time of participants was conducted, to examine whether participants required more time to respond during the more difficult trials. Reaction time was defined as the time between when the participant was shown all of the available choice stimuli and when they clicked on an individual stimulus to make a choice. Pearson's correlations were conducted between overall accuracy during each testing phase, and performance on Set C of the Raven's matrices, to evaluate whether success on one task was related to another.