Memory & Cognition (2020) 48:96–110
https://doi.org/10.3758/s13421-019-00964-y

Individual differences in relational reasoning

Maureen E. Gray[1] · Keith J. Holyoak[1]

Published online: 17 July 2019

Abstract
Various forms of relational processing have been linked to cognitive capacity measures, such as working memory and fluid intelligence. However, previous work has not established the extent to which different forms of relational processing reflect common factors, nor whether individual differences in cognitive style also contribute to variations in relational reasoning. The current study took an individual-differences approach to investigate the prerequisites for relational processing. In two studies, college students completed a battery of standardized tests of individual differences related to fluid intelligence and cognitive style, as well as a series of experimental tasks that require relational reasoning. Moderate correlations were obtained between relational processing and measures of cognitive capacity. Questionnaire measures of cognitive style generally did not improve predictions of relational processing beyond the influence of measures of cognitive capacity.

Keywords Relational reasoning · Individual differences · Cognitive capacity · Cognitive style

Introduction

Relational reasoning – inferential processes constrained by the relational roles that entities play rather than the specific features of those entities – is a hallmark of human cognition. Languages would be severely limited without prepositions and verbs that represent relations between things (e.g., *give* expresses an exchange of something between a giver and a recipient). Analogical reasoning, in which a familiar source domain is mapped to a less understood target domain that shares its relational structure, underlies the powerful ability to derive plausible inferences about a target based on a source analog. Analogical reasoning can be challenging when surface properties differ for entities that correspond across the analogs (e.g., Gick & Holyoak, 1980).

Given that many school subjects involve relational knowledge, understanding the cognitive underpinnings of such knowledge may help to improve education. Virtually all concepts in STEM fields are relational in nature (i.e., defined by shared relations rather than shared features). Furthermore, expertise in any domain requires rich knowledge of an interrelated set of concepts, many of which may themselves be relational in nature. Indeed, some researchers have argued that analogy is critical for creativity and innovation in technological fields (Goel, 1997). Goldwater and Schalk (2016) suggest that abstract relational schemas are prerequisites for knowledge transfer, which is arguably the end goal of education. In addition, recent research has shown that effective use of relational processing separates successful from unsuccessful students (McDaniel, Cahill, Robbins, & Wiener, 2014). A better understanding of relational processing, and why some students embrace it while others do not, could lead to improved educational outcomes.

A great deal of research indicates that adequate cognitive capacity (often characterized in terms of concepts such as working memory, inhibitory control, executive functioning, and/or fluid intelligence; see Ackerman, Beier, & Boyle, 2005) is necessary for relational processing (for a review, see Holyoak, 2012). For example, imposing a working-memory load causes college students to make fewer relational (and more featural) matches on a picture-mapping task (Waltz, Lau, Grewal, & Holyoak, 2000). Scores on the Ravens Progressive Matrices (RPM), a standard measure of fluid intelligence (Raven, 1938), have been shown to correlate positively with the probability of spontaneous analogical transfer in a problem-solving task (Kubricht, Lu, & Holyoak, 2017). Neuropsychological evidence links impaired prefrontal functioning with greatly diminished performance on analogy tasks (e.g., Kane & Engle, 2002; Krawczyk et al., 2008; Morrison et al., 2004). In one such study, participants completed an analogy task in which the correct answer was based on

✉ Maureen E. Gray
maureengray@ucla.edu

[1] Department of Psychology, University of California, 405 Hilgard Ave, Los Angeles, CA 90095-1563, USA

relational similarity. Individuals with frontal-variant frontotemporal lobar degeneration failed to inhibit a semantically related distractor, demonstrating the importance of interference control in relational processing (Krawczyk et al., 2008). A number of computational models of analogical reasoning emphasize the centrality of capacity constraints (e.g., Halford, Wilson, & Phillips, 1998; Hummel & Holyoak, 1997, 2003; Keane, 1997).

Relational reasoning may not depend entirely upon pure cognitive capacity, however. While fluid intelligence capabilities decline in older adulthood (Horn & Cattell, 1967), older adults are still able to complete relational tasks of lower relational complexity (Viskontas et al., 2004). Older adults do not simply lose their ability to reason upon reaching an advanced age, which suggests that some other constructs are at play. One other potential contributor to relational reasoning performance is crystalized intelligence, the counterpart to fluid intelligence that reflects reasoning based on prior knowledge. Indeed, in many situations, such as reaching a justified conclusion regarding the guilt or innocence of an individual based on information presented at a trial, reasoning critically depends on accumulated knowledge. Studies that have examined links between fluid intelligence and relational reasoning have seldom distinguished the potentially separable impact of crystalized intelligence.

While cognitive capacity and accumulated knowledge are likely to be important contributors to performance in tasks that require relational reasoning, other sources of individual differences may also play a role. In particular, some evidence implicates variations in cognitive style – differences in preferred thinking strategies – in performance on reasoning tasks (e.g., Stanovich & West, 1997). Several measures of cognitive style might be plausibly linked to relational processing. For example, the Need for Cognition (NFC) scale measures preferences for engaging in or avoiding analytic thinking (Cacioppo & Petty, 1982). Some past research has linked scores on the NFC to performance on the RPM, which is an inherently relational task (Day et al., 2007; Hill et al., 2013), as well as to performance on syllogistic reasoning problems (West, Toplak, & Stanovich, 2008). However, it remains unclear whether individuals' propensity to think analytically affects relational processing in other tasks that involve it, such as analogical reasoning. To the best of our knowledge, this paper is the first to explicitly examine the predictive power of cognitive style measures in the context of an extensive battery of relational processing tasks.

Another cognitive style measure that may be related to relational processing is the construct of Actively Open-minded Thinking (AOT; Baron, 1985). This scale captures an individual's propensity to avoid "myside" bias, which is the tendency to approach and process new information in such a way that already-held beliefs are strengthened (Baron, Scott, Fincher, & Metz, 2015). While this measure has been linked to some instances of reasoning (e.g., better performance on belief bias syllogisms; Baron et al., 2015), its possible relationships to other tasks involving relational processing have not been investigated.

As a final example, the Need for Cognitive Closure (NFCC; Kruglanski, 1989) scale measures an individual's desire to reach some answer on a given problem, regardless of whether or not that answer is correct. This desire manifests itself in a tendency to reach conclusions quickly and maintain them in an effort to reduce and avoid feelings of ambiguity. As in the case of the other cognitive style measures noted above, the NFCC has not been linked explicitly to relational reasoning tasks; however, the measure appears to be related to some characteristics that might influence reasoning, such as resistance to consideration of alternative hypotheses (Kruglanski & Mayseless, 1988) and a preference for simplified judgments (Van Hiel & Mervielde, 2003). This construct might plausibly relate to relational processing more generally.

One measure that arguably spans the gap between cognitive capacity and cognitive style is the Cognitive Reflection Test (CRT), developed by Frederick (2005). The CRT is a short test that measures what Frederick termed "cognitive reflection," which is "the ability or disposition to resist reporting the response that first comes to mind" (p. 35). The test consists of three word problems, each of which has an obvious, "intuitive" answer that springs to mind but is ultimately incorrect. To answer these problems correctly, people must inhibit the tendency to respond with the automatically generated incorrect answer and think more analytically. Frederick argued that the CRT is related to cognitive capacity, interference control, and cognitive style. He found the CRT to be correlated weakly with the NFC ($r = .22$), and moderately with three measures of cognitive capacity (the SAT, ACT, and Wonderlic Personnel Test; $r = .43$–$.46$). The CRT has also been shown to be weakly related ($r = .15$) to scores on the Stroop test, a common measure of inhibitory control (Toplak, West, & Stanovich, 2011). Neuropsychological research also suggests an inhibitory control component of the CRT. When participants' inhibitory-control capabilities are diminished by administering cathodal stimulation to the dorsolateral prefrontal cortex, they give more incorrect impulsive answers to CRT questions (Oldrati, Patricelli, Colombo, & Antonietti, 2016).

The present studies applied an individual-difference approach to investigate potentially separable components of relational reasoning. Our general tack was to examine a suite of tasks that appear to require relational processing, and to have participants also complete a battery of tests expected to measure aspects of cognitive capacity, inhibitory control, crystalized intelligence, and cognitive style. We sought (1) to determine which relational tasks seem to exhibit a shared pattern of relationships to measures of individual differences, and (2) to assess which types of individual differences predict performance on tasks requiring relational reasoning.

Study 1

Study 1 explored relationships among performance on a series of a tasks requiring relational processing and performance on a subset of the tests of individual difference reviewed above. The relational processing tasks included an analogical transfer problem (Gick & Holyoak, 1980), translating a statement into an algebraic expression (Martin & Bassok, 2005; Simon & Hayes, 1976), and a picture-mapping task (Markman & Gentner, 1993). These tasks, while heterogeneous in nature, were selected because they all involve some degree of relational reasoning and are used widely in the literature to study relational reasoning (e.g., Cushen & Wiley, 2018; Fisher, Borchert, & Bassok, 2011; Kubricht et al., 2017; Lewis & Mayer, 1987; Tohill & Holyoak, 2000; Vendetti, Wu, & Holyoak, 2014). Two of the selected tasks involve constructing an analogical mapping (the analogical transfer problem and the picture-mapping task), and all tasks involve consideration of relations between entities. The individual difference measures administered in Study 1 were Raven's Progressive Matrices (RPM; Arthur, Tubré, Paul, & Sanchez-Ku, 1999), the Cognitive Reflection Test (CRT; Frederick, 2005), and the Need for Cognition scale (NFC; Cacioppo & Petty, 1982).

Method

Participants

Participants were 202 undergraduate students (mean age = 20.1 years, 137 female) from the University of California, Los Angeles (UCLA) who received course credit for participating.

Measures

Each participant completed a series of individual difference measures, followed by experimental tasks likely to require relational processing.

Raven's Progressive Matrices (RPM) Participants completed a shortened, 12-item version of the RPM test, a common measure of fluid intelligence (Arthur et al., 1999). In this task, participants view a series of 3 × 3 grids with shapes in each cell except for the bottom right cell, which is blank. Systematic patterns are instantiated across the rows and down the columns of each matrix. From eight alternatives, participants choose which shape correctly completes the matrix by following the relational rules instantiated in the filled cells. This task is untimed and no feedback is given.

Need for Cognition (NFC) The Need for Cognition scale (Cacioppo & Petty, 1982), measures whether the individual enjoys engaging in analytic thinking. The shortened scale was used, which consists of 18 statements about processing preferences (e.g., "I would prefer complex to simple problems," or "Thinking is not my idea of fun"). Participants indicated how characteristic each statement is of themselves on a scale from 1 (extremely uncharacteristic) to 5 (extremely characteristic). Some items were reverse scored.

Cognitive Reflection Test (CRT) The Cognitive Reflection Test (Frederick, 2005), measures an individual's ability to inhibit an automatic response and engage in more effortful analytic thinking. This test consists of three problems, all of which have an "obvious" incorrect answer that immediately springs to mind. To answer these problems correctly, participants must inhibit these attractive automatically generated answers and instead engage in more effortful processing to compute the correct answer. The CRT is thought to tap many constructs, including cognitive capacity, inhibitory control, and cognitive style (Campitelli & Gerrans, 2014).

Analogical transfer In this task (Gick & Holyoak, 1980), participants read a story containing a source analog ("The General") and summarized it. Later, they were presented with the radiation problem (Duncker, 1945), which has an analogous "convergence" solution, and were prompted to solve it. After attempting to solve the problem without any prompt to use the source analog, participants were given a hint to think back to the source analog story and write down a solution that the story suggests. The total solution rate for the radiation problem was calculated based on convergence solutions generated either before or after the hint is given.

Algebra translation problem In this task (Martin & Bassok, 2005; Simon & Hayes, 1976), participants read the statement, "There are six times as many students as professors at this university," and were asked to translate it into an algebraic expression. This task requires mapping semantic elements of the verbal problem (number of students and of professors) onto elements of an algebraic equation. Success on this problem requires successfully avoiding a deceptively easy syntactic translation strategy, which would yield the incorrect expression $6S = P$. Producing the correct response, $S = 6P$, requires engaging in analytic processing and evaluating the qualitative relation between the number of students and of professors (the number of students would be expected to exceed the number of professors).

Picture-mapping task The final measure of relational processing employed was a picture-mapping task developed by Markman and Gentner (1993), with additional items added by Tohill and Holyoak (2000). In this task, participants were shown a series of picture pairs and asked to map one object from the top picture to an object in the bottom picture. The two pictures were displayed for 10 s, after which an object in the

top picture was visually highlighted (see Fig. 1). The participants then had to decide which object in the bottom picture "goes with" the highlighted object in the top picture. The expression "goes with" was kept purposefully vague: for each picture pair, the highlighted object could be mapped either on the basis of object attributes or on the basis of a shared relational role that each object fills. The dependent measure of interest was how many relational mappings a participant made out of 10 picture pairs.[1]

Procedure

Participants completed all tasks individually on a computer, using the keyboard to input responses. The tasks were ordered as follows: (1) Participants read the source analog for the analogical mapping task and summarized it. They then completed (2) the NFC scale, (3) the CRT, and (4) the RPM. Next, (5) participants were prompted to solve the radiation problem for the analogical mapping task; (6) they completed the algebra translation task; and (7) they completed the picture-mapping task. At the end of the study, participants were asked whether or not they had seen any of the tasks in the study previously, and if so to describe them. The study took 1 h to complete.

Results and discussion

Data from one participant who failed to follow experimental instructions were excluded, leaving a total of 201 participants for analysis. Data for specific tasks were excluded for several additional participants. Data from the CRT were excluded for six participants, and data from the analogical mapping problem were excluded for two others, because these participants expressed familiarity with the respective tasks. Data from the RPM were removed for one participant who failed to follow instructions and for eight additional participants whose log trial times fell below 2.5 standard deviations of the average log trial time for six or more trials, indicating low effort. Most seriously, the initial version of the instructions for the picture-mapping task proved confusing to participants, requiring us to modify the instructions. Data for the first 41 participants (who received the initial version) were excluded for this task. Data on this task were excluded for 12 additional participants because they gave five or more responses coded as "other," indicating a misunderstanding of the task. Open-ended responses were coded by two independent raters. Any disagreements were decided by a third party.

[1] A word classification task developed by Little and McDaniel (2015) was also administered. However, task performance was poor (fewer than 33% of participants successfully learned to classify the words). Consequently, data from this task were not analyzed and the task will not be discussed further. It was not included in the task battery for Study 2.

Solutions to the radiation problem were scored according to criteria adapted from previous research (Gick & Holyoak, 1980). If participants expressed at least two out of three critical ideas, they received full credit: (1) multiple radiation sources, (2) low intensity of rays, (3) arrangement of rays around the tumor with rays converging on the tumor. Responses were scored either as correct or incorrect (no partial credit was awarded). In addition, participants were scored as to whether they had solved the radiation problem spontaneously (without the hint) or after receiving the hint. The score could therefore be 0 (not solved), 1 (solved after hint), or 2 (solved without a hint). Inter-rater reliability was high for this task (Cohen's κ = .73).

Responses on the picture-mapping task were scored according to previously established criteria as featural, relational, or other (Markman & Gentner, 1993). The key dependent measure for this task was the number of relational mappings (out of 10 possible) that participants made. Inter-rater reliability on this task was high as well (Cohen's κ = .84).

Descriptive statistics

Raw means and standard deviations for the three key individual-difference measures and relational-processing measures are displayed in Table 1. These descriptive results show that performance on the analogical transfer task was poor. The spontaneous transfer rate in the current study (.09) was close to the solution rate found by Gick and Holyoak (1980) for participants who did not read a source analog (i.e., the control level). The total solution rate (.29) was much lower than that observed in the earlier study (.70). The poor performance in Study 1 may have been due to the extended time interval between presentation of the source analog and the target problem, coupled with interference from the demanding set of tasks that participants performed in between. As a result of the low spontaneous transfer rate, this task was recoded in a binary fashion as either solved (1) or not solved (0).

Correlational analyses

Prior to running analyses, each participant's score on each task was standardized. A relational composite measure was created by summing participants' standardized scores on each of the relational-processing measures (relational responses on the picture-mapping task, score on the algebra translation problem, and score on the analogical transfer task). We created this relational composite measure because there are theoretical reasons to believe that each of these tasks employs relational reasoning. Further, the pattern of correlations between the individual difference measures and each of the relational processing measures were comparable (see Table 2), suggesting that there may be some common process underlying each of the tasks. The relational processing measures did not correlate

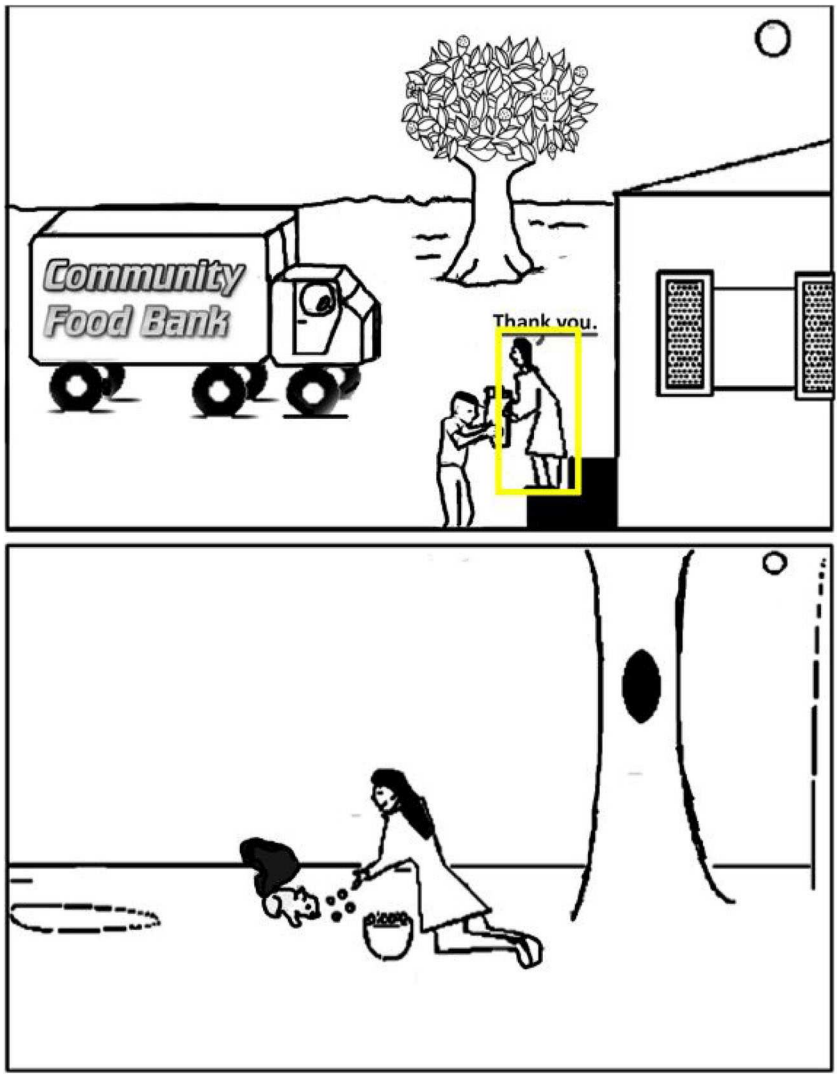

Fig. 1 Example picture from the Picture Mapping task used in Studies 1 and 2. The highlighted object from the top picture (the woman) could be mapped to the woman from the bottom picture (featural mapping) or to the squirrel (relational mapping). Adapted from Markman and Gentner (1993)

strongly with one another, perhaps due to the fact that two of the measures were binary rather than continuous, and that performance on the radiation problem was poor.

Inter-task correlations are presented in Table 2. In all correlational analyses, missing data were handled using pairwise deletion. Relationships between continuous variables were assessed with Pearson's correlations, while relationships between one continuous and one binary variable were assessed with point-biserial correlations. Finally, the phi coefficient was used to measure the association between two binary variables. Several interrelationships among the individual difference measures are

Table 1 Descriptive statistics (Study 1)

	N	Maximum Possible score	Mean	Std. deviation
RPM	192	12	7.05	2.81
CRT	195	3	1.05	1.11
NFC	201	90	60.63	10.49
Radiation problem	198	1	.30	.46
Algebra problem	199	1	.53	.50
Picture mapping	138	10	6.68	2.72

RPM Raven's Progressive Matrices, *CRT* Cognitive Reflection Test, *NFC* Need for Cognition

Table 2 Inter-task correlations (Study 1)

		1	2	3	4	5	6	7
1	RPM	1						
2	CRT	.48**	1					
3	NFC	.16*	.24**	1				
4	Radiation problem	.26**	.30**	.13	1			
5	Algebra problem	.34**	.36**	.23**	.18*	1		
6	Picture mapping	.37**	.33**	.06	.15	.23**	1	
7	Composite	.45**	.46**	.22*	.65**	.68**	. 67**	1

RPM Raven's Progressive Matrices, *CRT* Cognitive Reflection Test, *NFC* Need for Cognition

** denotes $p < .01$, * denotes $p < .05$

apparent. The moderate correlation between RPM and CRT ($r = .48$, $p < .001$) is somewhat stronger than correlations noted in previous studies, which have found these two measures to be correlated at about .3 (e.g., Brañas-Garza, García-Muñoz, & Hernán-González, 2012; Hanaki, Jacquement, Luchini, & Zylbersztejn, 2016). The weak relationship between RPM and NFC ($r = .16$, $p = .02$) is similar to that found in previous studies (e.g., Hill et al., 2013). Finally, the relationship between NFC and CRT ($r = .24$, $p = .001$) is similar to correlations found in previous studies, supporting the hypothesis that the CRT is sensitive to both capacity and style components (e.g., Pennycook, Cheyne, Koehler, & Fugelsang, 2016).

Table 2 also shows the pattern of correlations among the individual-difference measures and the relational-processing measures. In line with expectations from previous research (Kubricht et al., 2017), RPM scores correlated significantly with each of the relational reasoning measures, suggesting that cognitive capacity contributes to performance in these tasks. The CRT also correlated moderately with each of the relational reasoning measures. These correlations suggest the involvement of the constructs that contribute to performance on the CRT, including cognitive capacity, inhibitory control, and cognitive style. The pattern of correlations with NFC differed slightly between the three measures of relational reasoning. A significant relationship was observed between the NFC and the algebra problem, ($r_{pb} = .23$, $p < .001$), but not among the other two relational processing measures. The algebra problem may depend more upon cognitive style than the other two measures, although the relationships observed between the CRT (a behavioral measure of cognitive style) and each of the relational processing measures suggest that cognitive style may play a role more broadly.

Next, correlations between the individual-difference variables and the relational composite were examined. The relational composite measure was correlated moderately with RPM ($r = .45$, $p < .001$) and CRT ($r = .46$, $p < .001$), and weakly with NFC ($r = .22$, $p = .015$). This pattern suggests that relational processing was influenced by fluid intelligence and the constructs tapped by the CRT, which include cognitive capacity, style, and inhibitory control (Campitelli & Gerrans, 2014).

To further tease apart the independent contributions of each individual difference variable to relational processing performance, a stepwise multiple regression was run, predicting the relational composite from the three individual difference measures. Stepwise regression was utilized due to the exploratory nature of the analysis. We did not have *a priori* reasons to enter the predictors in any particular order, and wanted to account for potential shared variability among the predictors. The results of this analysis are shown in Table 3. In this analysis, participants with missing data were excluded on a listwise basis, leaving a total of 118 participants in the final analysis. The first predictor to enter the model was standardized score on the CRT, followed by standardized score on RPM. NFC did not contribute any unique predictive power after accounting for the CRT and RPM. Overall, this analysis supports the results of the correlational analysis, showing that RPM and the CRT each contributed unique predictive power with respect to relational processing. The impact of cognitive style was mediated by the behavioral CRT score but not by the self-assessed NFC test.

In sum, the findings of Study 1 suggested that there may be multiple distinct individual differences that support relational processing. Multiple regression analyses demonstrated that cognitive capacity contributes to relational reasoning performance. Although self-assessed measures of cognitive style were not predictive, style assessed in a behavioral manner contributed to relational reasoning.

Table 3 Multiple regression analyses predicting relational composite score (Study 1)

Predictor	ΔR^2	β	t
Step 1	.208**		
CRT		.456	5.53**
RPM		.274[a]	3.14**
NFC		.097[a]	1.20
Step 2	.059**		
CRT		.315	3.40**
RPM		.281	3.03**
NFC		.103[b]	1.31

Note. All measures are standardized.

RPM Raven's Progressive Matrices, *CRT* Cognitive Reflection Test, *NFC* Need for Cognition

a Predictors in the model: CRT

b Predictors in the model: CRT, RPM

**denotes $p < .01$

Study 2

Study 2 was performed to replicate the major findings of Study 1, while adding additional measures. The new measures included a new task that may involve relational processing, a measure of visual working memory span (Foster et al., 2015), a measure of crystalized verbal intelligence (Stamenković, Ichien, & Holyoak, 2019), and two additional self-report measures of cognitive style, the Need for Cognitive Closure scale (NFCC; Kruglanski, 1989) and the Actively Open-Minded Thinking scale (AOT; Baron et al., 2015).

Method

Participants

Participants were 231 UCLA undergraduate students (mean age = 20.3 years, 177 female) who received course credit for participating.

Measures

All the tasks assessed in Study 1 were also used in Study 2. Here we describe additional measures that were added to the test battery in Study 2.

Symmetry span The symmetry span task is a visuospatial complex span test designed to measure working-memory capacity (Foster et al., 2015). In this task, participants are presented with a 4 × 4 grid of cells. On each trial, a certain number of the cells are highlighted, one at a time. The participant's task is to correctly recall which cells were highlighted in which order. In between the presentation of each to-be-remembered highlighted cell, the participant was shown a figure and had to decide whether or not the figure was left-right symmetrical. After the presentation period ended, the participant was presented with a blank grid. The participant inputted his or her responses by clicking on the cells in the order in which they recalled seeing the cells highlighted. All participants completed three trials at each span length from 2 to 7 for a total of 21 trials. In addition, participants received feedback at the conclusion of each trial, informing them whether they had recalled the cells correctly or not. Scores were calculated by computing the average proportion of correctly recalled grids across all trials. Trials were weighted equally, so answering incorrectly at lower span lengths is more detrimental to the final score.

Semantic Similarities Test (SST) The SST is a short test designed to assess verbal crystalized intelligence (Stamenković et al., 2019) by asking participants to generate similarities for a given pair of concepts. Stamenković et al. (Study 3) found that scores on the SST were strongly correlated (r = .67) with performance on the WAIS-III Vocabulary subtest, a standard measure of crystalized verbal knowledge, whereas the test was correlated only moderately (r = .39) with the RPM, which is considered a measure of fluid intelligence. The SST consists of 20 word pairs ordered from easy (e.g., *orange-ball*) to hard (e.g., *tavern-church*). For each pair, the participant answered the question, "How are these two concepts similar?" (e.g., for *orange-ball* both are spheres, and for *tavern-church* both are places of gathering). The instructions for the SST included one example (*chair-sofa*) and a possible answer (both are types of furniture). Participants' scores were calculated using an answer key developed by Stamenković et al. Each item could be fully correct (2 points), partially correct (1 point), or incorrect (0 points).

Need for Cognitive Closure (NFCC) We supplemented the Need for Cognition scale with two additional scales that were designed to measure slightly different aspects of cognitive style. The Need for Cognitive Closure scale (Kruglanski, 1989) measures an individual's desire for a definite answer on some topic or problem in an effort to avoid ambiguity and confusion, regardless of whether that answer is correct or not. In the current study, a shortened scale devised by Roets and Van Hiel (2011) was used. The scale contains 15 items (e.g., "When I am confronted with a problem, I'm dying to reach a solution very quickly," or "I don't like situations that are uncertain").

Actively Open-minded Thinking (AOT) The measure of Actively Open-Minded Thinking (Baron, 1985) captures the disposition to weigh new evidence against a held belief (or not). The current study used a shortened measure consisting of eight items (e.g., "Allowing oneself to be convinced by an opposing argument is a sign of good character"). The shortened scale was developed by Baron, Scott, Fincher, and Metz (2015).

The three cognitive style scales (NFC, NFCC, and AOT) were intermixed and items were randomized once (i.e., each participant answered the items in the same randomized order). Participants were instructed to read each statement and decide how much they agree or disagree with each according to their beliefs and experiences. A 7-point Likert scale was used, where 1 represented "strongly disagree" and 7 represented "strongly agree."

Perceptual mapping task We administered an additional measure of relational processing adapted from a study by Goldstone, Medin, and Gentner (1991, Experiment 1). In this task, participants had to choose which of two possible figures was most similar to a given target figure. The target figures were created based on the stimuli provided in Goldstone et al., and the two response options were created to highlight either attributional or relational similarity to the target figure. There

were seven trials, and on each trial participants saw the target figure at the top of the screen with the two response options displayed below (see Fig. 2). Trial order and left/right presentation of response options (attributional or relational) were randomized, and participants recorded their responses by pressing the "a" key to indicate the left response option was most similar to the target, and pressing the "b" key to indicate the right response option was most similar to the target.

Procedure

Tasks were completed individually on a computer. The task order was as follows: (1) Participants completed the symmetry span task, followed by (2) the CRT, and (3) the SST. Next, participants took the (4) cognitive style composite scale, followed by (5) RPM. Then, participants completed (6) the perceptual mapping task, (7) read the source story for the analogical transfer problem, and (8) completed the algebra translation problem. Next, participants were prompted to (9) solve the radiation problem (first without a hint, then with a hint prompting them to think back to the source story), and (10) complete the picture mapping task. Note that in an effort to increase spontaneous analogical transfer rates, in Study 2 we moved the source analog and the target radiation problem closer together (separated only by the algebra translation problem). Finally, participants were asked whether or not they had seen any of the tasks in the study previously, and if so to describe them. The study took 1 h to complete.

Results and discussion

Data from one participant who did not follow experimental instructions were excluded, leaving a total of 230 participants for analysis. In addition, data for specific tasks were excluded for several additional participants. Data from the CRT were excluded for 14 participants, and data from the analogical transfer problem were excluded for 13 others, because these participants expressed familiarity with the respective tasks. Eight participants were excluded from the symmetry span task because their symmetry judgment accuracy fell below 85%, and two participants were excluded from RPM because the log of their trial response times (RTs) fell below 2.5 standard deviations of the mean log trial RT for six or more trials. Two participants were excluded from the SST because they scored below 12/40 points, which was identified by Stamenković et al. (2019) as the cutoff point. Finally, 22 participants were excluded from the picture-mapping task because they gave five or more responses coded as "other," indicating a misunderstanding of the task.

Coding for open-ended tasks was completed by two independent raters following the same criteria outlined in Study 1. Inter-rater agreement was comparable to that of Study 1 for both the analogical transfer problem and the picture-mapping task (Cohen's κ = .85 and .83, respectively). Any disagreements were decided by a third party.

Descriptive statistics

Raw means and standard deviations for each task are displayed in Table 4. As in Study 1 (and despite reducing the temporal separation between presentation of the story and radiation problem), spontaneous analogical transfer occurred very infrequently (.12 of participants), and the total solution rate of .33 was considerably lower than that observed by Gick and Holyoak (1980). Presumably analogical transfer was difficult in the context of the demanding overall battery of tasks. As in Study 1, the analogical transfer task was recoded into a binary variable where participants received a score of 1 if they solved the radiation problem or a score of 0 if they failed to do so.

In traditional scoring of the NFCC, higher scores indicate a greater need for cognitive closure. In the present study, scores were inverted to match the structure of the other cognitive style scales, so that higher scores correspond to a greater

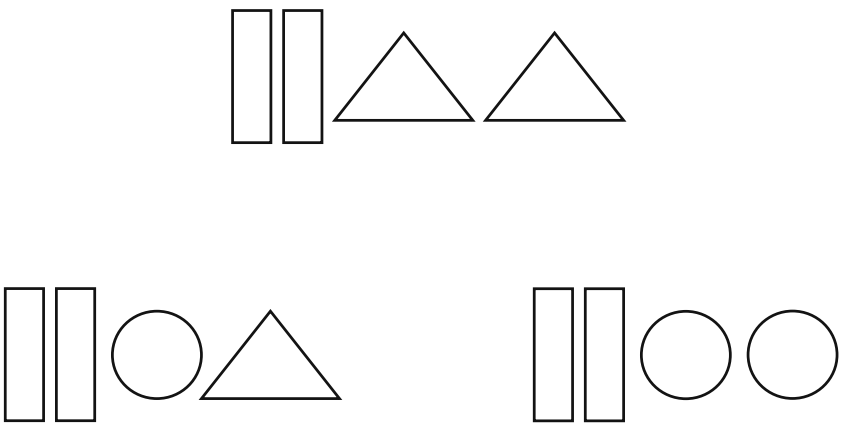

Fig. 2 Example stimulus from the Perceptual Mapping task used in Study 2. Participants could choose the left figure (showing a preference for featural similarity) or the right figure (showing a preference for relational similarity). Adapted from Goldstone, Medin, and Gentner (1991)

Table 4 Descriptive statistics (Study 2)

	N	Maximum Possible Score	Mean	Std. Deviation
RPM	228	12	6.50	2.92
Symmetry span	222	1	.63	.18
CRT	217	3	0.99	1.03
SST	229	40	26.30	4.30
NFC	229	126	55.00	12.73
AOT	229	56	38.51	6.57
NFCC	229	105	80.48	13.47
Ray problem	207	1	.33	.47
Algebra problem	224	1	.50	.50
Picture mapping	184	10	6.22	2.56
Perceptual mapping	227	7	2.74	2.81

RPM Raven's Progressive Matrices, *CRT* Cognitive Reflection Test, *SST* Semantic Similarities Test, *NFC* Need for Cognition, *AOT* Actively Open-minded Thinking, *NFCC* Need for Cognitive Closure

tolerance of uncertainty and ambiguity, and lower scores correspond to a greater need for cognitive closure.

As shown in Table 4, on average participants made two relational matches out of seven possible matches on the perceptual mapping task, demonstrating low levels of relational responding overall. This pattern is comparable to that reported by Goldstone et al. (1991), who observed that when given the choice between these two particular kinds of figures, participants tended to make attributional matches. However, considering only mean performance on this task may be misleading. A histogram revealed that the distribution of responses on this task appeared to be bimodal. One group of participants (n = 88) made only attributional matches, while another group (n = 53) made almost only relational matches. The basis for this bimodal responding is uncertain, but two general possibilities should be considered. First, individual participants may only perceive one type of similarity. Alternatively, participants may perceive both relational and attributional similarity, and focus on the one that they prefer (since the instructions do not favor one type of response over the other). Thus, the apparent individual differences in choice may arise either at the perceptual level of processing or at a later decision stage. Our data do not discriminate between these two possibilities, but we note the potential for future investigations that could use this task to investigate the basis for individual differences in processing relational versus attributional similarity.

Correlational analyses

As in Study 1, all scores on all tasks were standardized prior to analysis. Again, as in Study 1, missing data were handled with pairwise deletion. The pattern of correlations among the relational processing measures observed in Study 2 was weaker than that observed in Study 1 (see Table 5), though the three measures used in Study 1 showed similar relationships with the individual difference measures. As a result of the overall low levels of relational responding in the perceptual mapping task, this task was excluded from the composite relational measure. Thus, as in Study 1, the relational composite measure was constructed by summing together participants' standardized scores on three measures (relational responses on the picture-mapping task, score on the algebra translation problem, and score on the analogical transfer task).

Table 5 also shows correlations among all individual difference measures and the relational composite. Several of the relationships observed in Study 1 were also observed in Study 2. The moderate correlation between RPM and CRT (r = .49, p < .001) was replicated, but the relationship between NFC and RPM was not. However, the RPM correlated weakly with both AOT (r = .21, p = .001) and NFCC (r = .21, p = .001). The correlation between the CRT and NFC in Study 2 (r = .14, p = .03) was slightly weaker than that observed in Study 1 (r = .24, p = .001).

Correlations among the new individual-difference measures added in Study 2 were examined. The moderate correlation between RPM and symmetry span (r = .42, p < .001) is comparable to those observed in previous studies (e.g., Foster et al., 2015), and the moderate correlation between NFC and AOT (r = .30, p < .001) is also comparable to previous findings (e.g., Haran, Ritov, & Mellers, 2013; West et al., 2008). Contrary to expectations, the cognitive style measures did not correlate consistently with one another. While a moderate correlation was observed between AOT and NFCC (r = .38, p < .001), the NFC was only weakly correlated with NFCC (r = .17, p = .011). Previous studies have found the correlation between NFC and NFCC to be closer to .3 (e.g., Petty & Jarvis, 1996). Moreover, the correlation between NFC and AOT was weak and negative (r = -.13, p = .05). In contrast,

Table 5 Inter-task correlations (Study 2)

		1	2	3	4	5	6	7	8	9	10	11
1	RPM	1										
2	Symmetry span	.40**	1									
3	CRT	.54**	.38**	1								
4	SST	.25**	.14*	.22**	1							
5	NFC	.21**	.02	.14*	.10	1						
6	AOT	.22**	.10	.22*	.30**	.38**	1					
7	NFCC	.01	.07	.01	.07	-.16*	.13*	1				
8	Ray problem	.21**	.15*	.25**	.17*	.02	.09	.02	1			
9	Algebra problem	.19**	.17*	.37**	.17*	-.05	.08	-.04	.09	1		
10	Picture mapping	.26**	.22**	.22**	.27**	-.10	.25**	.10	.20*	.08	1	
11	Perceptual mapping	-.06	-.07	-.04	-.03	.10	.01	.17*	-.12	-.22**	.03	1
12	Composite	.31**	.23**	.39**	.31**	-.11	.20**	-.01	.67**	.59**	.66**	-.12

RPM Raven's Progressive Matrices, *CRT* Cognitive Reflection Test, *SST* Semantic Similarities Test, *NFC* Need for Cognition, *AOT* Actively Open-minded Thinking, *NFCC* Need for Cognitive Closure

** denotes $p < .01$, * denotes $p < .05$

the positive correlation observed between RPM and the SST ($r = .25$, $p < .001$) was consistent with that observed in previous research (Stamenković et al., 2019), as well as with theoretical expectations (in that fluid intelligence and crystalized intelligence are both expected to load on general intelligence).

Finally, correlations between the individual difference measures and the relational processing tasks were examined to shed light on which of the measured individual differences play a role in relational processing. As in Study 1, the relational composite measure was correlated with RPM ($r = .31$, $p < .001$). The other measure of "pure" cognitive capacity, symmetry span, was also correlated with the relational processing composite ($r = .23$, $p = .010$). A moderate correlation was observed between the relational composite and the CRT ($r = .39$, $p < .001$), suggesting the contribution of the constructs involved in the CRT, which include cognitive capacity, inhibitory control, and cognitive style (Campitelli & Gerrans, 2014). In addition, crystalized intelligence as assessed by the SST was modestly correlated with the relational composite ($r = .30$, $p < .001$), suggesting that in addition to raw cognitive power, accumulated verbal knowledge also contributes to relational processing.

In contrast to the indirect assessment of cognitive style tapped by the CRT, correlations between the self-report cognitive style measures and the relational composite were not consistently obtained in Study 2. The correlations between NFC and NFCC and the relational composite were near 0, and the relationship between AOT and the relational composite was weak ($r = .20$, $p = .009$). These correlational analyses generally replicate the findings of Study 1, supporting the hypothesis that relational processing relies most prominently on cognitive capacity and accumulated knowledge, while the contribution of "pure" cognitive style as assessed by self-report measures is less clear. However, the consistent relationship observed between the CRT and the relational composite suggests that less explicit measures of cognitive style may be related to relational reasoning.

To investigate the independent contribution of each of the individual difference constructs to relational processing, we ran a stepwise multiple regression predicting performance on the relational composite from each of the individual difference measures. As in Study 1, stepwise regression was selected over other model selection procedures because we had no *a priori* reasons to enter predictors in a particular order and wanted to account for any shared variability among predictors. Participants with any missing data were excluded on a listwise basis, leaving a total of 153 participants in the analysis. As shown in Table 6, standardized score on the CRT was the first predictor to enter the model, followed by standardized score on the SST. As in Study 1, none of the self-report measures of cognitive style contributed any unique predictive variance to the model. Unlike Study 1, however, the RPM did not contribute any additional predictive variance after accounting for the CRT and the SST. To further investigate the relationship between the CRT, RPM, and relational processing, an additional stepwise multiple regression analysis was run predicting performance on the relational composite from CRT, RPM, and NFC using observations from both studies. After excluding participants with missing data in a listwise fashion, the total number of observations in this analysis was 277. The results of this analysis, shown in Table 7, support the findings of Study 1. The CRT entered the model first, followed by the RPM. This analysis supports the hypothesis that the CRT and RPM, while related, nonetheless assess separable processes involved in relational reasoning.

Table 6 Multiple regression analyses predicting relational composite score (Study 2)

Predictor	ΔR^2	β	t
Step 1	.122**		
CRT		.358	4.71**
RPM		.133[a]	1.56
WM		.068[a]	0.81
SST		.195[a]	2.56*
AOT		.157[a]	2.03*
NFC		-.154[a]	-2.04*
NFCC		-.053[a]	-0.70
Step 2	.036*		
CRT		.315	4.12**
SST		.195	2.56*
RPM		.105[b]	1.26
WM		.054[b]	0.65
AOT		.122[b]	0.12
NFC		-.135[b]	-1.82
NFCC		-.074[b]	-0.98

Note. All measures are standardized

RPM Raven's Progressive Matrices, *CRT* Cognitive Reflection Test, *SST* Semantic Similarities Test, *NFC* Need for Cognition, *AOT* Actively Open-minded Thinking, *NFCC* Need for Cognitive Closure, *WM* working memory

a Predictors in the model: CRT

b Predictors in the model: CRT, SST

** denotes $p < .01$, * denotes $p < .05$

The multiple regression results in Study 2 generally confirm and extend the results obtained in Study 1. Cognitive capacity, inhibitory control, crystalized intelligence, and a behavioral assessment of cognitive style contributed predictive variance to relational processing, whereas cognitive style as assessed by standard self-report questionnaires (NFC, NFCC, AOT) was not observed to have an independent effect. In a multiple regression analysis combining participants from both studies, the CRT and RPM added separate predictive power to relational reasoning performance. Given that the CRT accounts for a significant amount of variation in relational processing after accounting for cognitive capacity as assessed by RPM, it seems that this measure uniquely captures an inhibitory control component of cognitive capacity and some aspect of cognitive style as it relates to relational processing performance.

Table 7 Multiple regression analyses predicting relational composite score (Studies 1 and 2)

Predictor	ΔR^2	β	t
Step 1	.172**		
CRT		.415	7.57**
RPM		.206[a]	3.35**
NFC		-.025[a]	-0.45
Step 2	.032**		
CRT		.315[b]	5.10**
RPM		.206[b]	3.35**
NFC		-.025[b]	-0.46

Note. All measures are standardized

RPM Raven's Progressive Matrices, *CRT* Cognitive Reflection Test, *NFC* Need for Cognition

a Predictors in the model: CRT

b Predictors in the model: CRT, RPM

**denotes $p < .01$

General discussion

Summary

The present study applied an individual differences approach to investigate component processes underlying relational processing in tasks related to analogical reasoning. In two studies, large samples of college students completed a battery of relational tasks as well as a set of tests designed to assess cognitive capacity, inhibitory control, crystalized intelligence, and cognitive style. The relational tasks were used to construct a composite measure of relational processing. This composite measure was based on analogical transfer in a verbal problem-solving task (Gick & Holyoak, 1980), an algebra task requiring translation from a verbal problem to an equation (Simon & Hayes, 1976), and a task requiring mapping objects between two visual scenes (Markman & Gentner, 1993). Regression analyses were performed to identify measures that made separable contributions to prediction of this relational composite. In Study 1, scores on the Cognitive Reflection Task (CRT) and the Ravens Progressive Matrices (RPM) proved to be effective predictors, whereas score on the Need for Cognition test (NFC, a self-report measure of cognitive style) did not. Study 2 included a more extensive battery of individual difference measures, including the Semantic Similarities Test (SST) as a measure of crystalized intelligence, and two additional measures of cognitive style: Need for Cognitive Closure (NFCC) and Actively Open-minded Thinking (AOT). An overall regression analysis combining the data from the two studies indicated that the CRT and RPM contributed separable predictive power. In addition, Study 2 revealed that the SST also made a separable contribution, suggesting that verbal semantic knowledge makes a unique contribution to aspects of relational reasoning. Each of the three self-report measures of cognitive style yielded weak correlations with the measures of cognitive capacity, but none of these cognitive style tests improved prediction of

relational processing after statistically removing the influence of the other measures.

Implications for component processes underlying relational reasoning

The relational processing tasks selected for use in these studies were heterogeneous in nature and correlated with one another only weakly. However, each of the selected tasks involves relational processing in some form, and a consistent pattern of correlations was observed between the individual difference measures and the each of the relational tasks. The fact that the tasks were so different at the surface level yet elicited a common pattern of correlations extends the generality of the current findings.

One previous study that examined the genetic basis of complex relational processing found correlations among relational tasks ranging from .40 to .56 (Hansell et al., 2015), which were significantly larger than those observed in the current studies. However, Hansell et al. specifically chose relational tasks in the context of relational complexity theory (Halford et al., 1998) with a goal of quantifying the single construct of relational complexity across different domains (e.g., verbal comprehension, deductive reasoning). In contrast, we chose relational tasks commonly used in research on relational reasoning, but without the goal of defining a specific underlying construct.

The individual differences in relational reasoning observed in the present study help to characterize the component processes involved in this type of reasoning. Given what prior research has indicated about the nature of the various measures of individual differences employed in the present studies, it is clear that what is broadly considered fluid intelligence, or executive functions, plays a major role. Performance on the CRT is believed to reflect the ability to inhibit the impulse to accept an "obvious" answer uncritically, and to think flexibly (shifting strategies as needed). The RPM appears to assess the ability to form and maintain goals and subgoals in working memory, as well as the ability to infer relations and to use them to generate inferences. Although performance on these two tests was correlated, the aggregated data from the two studies indicated that each made an independent contribution to prediction of success on the relational composite measure. Notably, in Study 2 a relatively pure measure of working-memory capacity (symmetry span) failed to make an independent contribution after accounting for the impact of the CRT and RPM. These results are consistent with the view that working memory and fluid intelligence, though closely related, are not identical constructs (Ackerman et al., 2005).

Our findings are consistent with those of previous studies that identified a link between relational processing and fluid intelligence (Kubricht et al., 2017; Vendetti et al., 2014). The present work also supports previous findings linking inhibitory control to relational processing (e.g., Cho et al., 2010; Krawczyk et al., 2008). Previous studies have found the CRT to be positively correlated with performance on various decision-making tasks (e.g., Lesage, Navarrete, & de Neys, 2013; Toplak et al., 2011), and with rule transfer in a causal learning paradigm (Don, Goldwater, Otto, & Livesey, 2016), while it is negatively correlated with trust in intuition (Pennycook et al., 2016). Given that a link between relational processing and the CRT has now been established, the relevance of relational processing to each of these tasks should be considered. For example, an individual who is better at processing relations might be more likely to consider the relations between variables in a conjunctive probability problem.

In Study 2 we found that the SST, a measure of crystalized intelligence based on verbal semantic knowledge, is also a potent predictor of relational processing. It is noteworthy that although many studies of cognitive individual differences have included the RPM and other measures of fluid intelligence and executive functions, almost none have included tests that assess semantic knowledge. It seems likely that a rich store of semantic relations in long-term memory can reduce the burden on working memory during relational reasoning tasks involving strong semantic content (e.g., solving the radiation problem using a convergence analog). It is noteworthy that the SST has been shown to predict metaphor comprehension, and in fact appears to be a stronger predictor than the RPM for relatively simple metaphors (Stamenković et al., 2019). Each of the three tasks comprising our relational composite involved active manipulation of semantic knowledge (whether conveyed verbally or in meaningful visual scenes). The importance of crystalized verbal intelligence as demonstrated in the present study is consistent with computational models of relational processing that treat knowledge of semantic relations as a core constraint (e.g., Doumas, Morrison, & Richland, 2018; Holyoak & Thagard, 1989; Lu, Chen, & Holyoak, 2012: Lu, Wu, & Holyoak, 2019).

In contrast to the clear predictive power of core measures of fluid and crystalized intelligence, we found no compelling evidence that questionnaire measures of cognitive style can add additional predictive power. In general, the cognitive style measures showed weak correlations with the capacity-related measures, but did not contribute separately to prediction of success in relational reasoning. However, the self-report measures of cognitive style used in the current studies may underestimate the contribution of cognitive style to relational reasoning, and the limited selection of cognitive style measures used in the current project necessitates caution in interpreting null results. Indeed, previous research shows that individuals with low analytic thinking abilities (assessed through performance on the CRT) demonstrate a systematic miscalibration of perceived Need for Cognition: these participants reliably

overreport their own Need for Cognition when their performance on a behavioral measure suggests otherwise (Pennycook, Ross, Koehler, & Fugelsang, 2017). Such findings suggest that self-report measures of cognitive style do not necessarily reflect behavioral tendencies. To clarify the true contribution of cognitive style to relational reasoning, future work should include a more extensive battery of behavioral cognitive style measures, such as an expanded CRT (Thomson & Oppenheimer, 2016), base-rate problems, and heuristics and biases problems (e.g., Stanovich & West, 2008).

One previous study administered the NFC along with some behavioral indices of cognitive style (including the CRT) and measures of cognitive capacity, as well as a test of verbal analogical reasoning (Barr, Pennycook, Stolz, & Fugelsang, 2015). A modest correlation was observed between NFC and accuracy on validation of cross-domain analogies ($r = .36$). In a multiple regression analysis, the behavioral indices of cognitive style and cognitive ability measures independently predicted performance on a composite measure consisting of performance on cross-domain analogies and the Remote Associates Test (a common measure of creativity). This regression analysis did not include questionnaire measures of cognitive style. The results reported by Barr et al. are consistent with the present findings in showing that behavioral measures of cognitive style and cognitive capacity are related to analogical reasoning tasks.

We emphasize that the present null findings do not imply that cognitive style has no impact on reasoning. In particular, it has been suggested that the CRT in part reflects thinking dispositions (Campitelli & Gerrans, 2014; Pennycook et al., 2014; Toplak et al., 2011). Given that the CRT was shown to add additional predictive power separately from the influence of cognitive capacity measures, differences in thinking dispositions may have contributed to the predictive potency of the CRT. It is noteworthy that the CRT does not directly signal the relevance of cognitive style to a person taking the test, in contrast to the direct style measures, which are based on questionnaires that explicitly query thinking dispositions. Although these questionnaire measures of cognitive style typically correlate with each other and with a variety of personality scales (e.g., self-consciousness, dogmatism, and introspectiveness; Cacioppo, Petty, Feinstein, & Jarvis, 1996), they may not independently predict relational reasoning performance. The present results imply that the impact of cognitive style measures on reasoning should be assessed while taking into account the impact of correlated capacity variables (e.g., Kokis, Macpherson, Toplak, West, & Stanovich, 2002; Stanovich & West, 1997). In addition, behavioral measures of cognitive style should be used whenever possible (Pennycook et al., 2017).

Although in the present study we did not find any clear links between the self-report cognitive style measures and relational reasoning, such links have been found for some other reasoning tasks (e.g., Griffin, Wiley, Britt, & Salas, 2012; Stanovich & West, 1997). Cognitive style measures have been linked to probabilistic reasoning (Kokis et al., 2002), syllogistic reasoning problems and belief bias (West et al., 2008; Macpherson & Stanovich, 2007), argument evaluation (Stanovich & West, 1997), and a myriad of heuristics and biases tasks (Toplak et al., 2011). It is noteworthy that those reasoning tasks for which performance can be predicted by questionnaire measures of cognitive style have largely been drawn from the literature on heuristics and biases, whereas those not predicted by the questionnaire measures (i.e., those used in the present paper) have been largely drawn from the literature on analogical reasoning. Relational reasoning is likely not a unitary construct. It remains an open question what factors distinguish the types of relational reasoning that are or are not predicted by different measures of cognitive style.

Approaches to improving relational reasoning

Given the strong contribution of executive function to successful relational processing, interventions to improve relational reasoning should focus on bolstering the cognitive capacity of the reasoner, or on lessening their cognitive load. Much work in the field of educational psychology has focused on the benefit of lessening the cognitive burden on the learner (Sweller, 2011). For example, studies have shown that for lower-skill students in particular, lessening the burden on cognitive capacity by first presenting a worked example of a problem before asking the student to generate their own solution leads to superior learning outcomes (e.g., Barbieri & Booth, 2016; Reisslein, Atkinson, Seeling, & Reisslein, 2006). Previous studies of relational reasoning offer examples of the kinds of interventions that may help individuals with relatively low cognitive capacity. Kubricht et al. (2017) found that supplying an animated diagram along with the source analog improved analogical transfer performance for individuals with lower scores on the RPM. Vendetti et al. (2014) showed that generating solutions to semantically distant analogies induced a general relational set, which in turn increased the number of relational matches made on an unrelated picture-mapping task (see also Andrews & Bohadana, 2018). Moreover, induction of a relational set reduced the association between performance on the mapping task and score on the RPM. Given the ubiquity of relational processing in math and science education (Goldwater & Schalk, 2016), improving relational reasoning is an important goal in the overall effort to improve educational outcomes.

Acknowledgements This research was supported by NSF Grant BCS-1827374. A preliminary report of part of this study was presented at the 40th Annual Conference of the Cognitive Science Society (Madison WI, July 2018). We thank Daniela Rojas and Rina Fujikawa for assistance with testing and data analyses, and Dušan Stamenković for advice on the Semantic Similarities Test.

Open Practices Statement Neither of the studies reported in this article was formally preregistered. Neither the data nor the materials have been made available on a permanent third-party archive, but requests for the data or materials can be sent via email to the lead author at maureengray@ucla.edu.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*, 30–60.

Andrews, G., & Bohadana, G. R. (2018). Does solving distant analogies facilitate relational integration? *Journal of Cognitive Psychology*, *30*(3), 270-280.

Arthur, W., Tubré, T. C., Paul, D. S., & Sanchez-Ku, M. L. (1999). College-sample psychometric and normative data on a short form of the Raven Advanced Progressive Matrices Test. *Journal of Psychoeducational Assessment*, *17*(4), 354–361.

Barbieri, C., & Booth, J. L. (2016). Support for struggling students in algebra: Contributions of incorrect worked examples. *Learning and Individual Differences, 48*, 36-44. https://doi.org/10.1016/j.lindif.2016.04.001

Baron, J. (1985). Rationality and intelligence. New York: Cambridge University Press.1985-98431-000

Baron, J., Scott, S., Fincher, K., & Emlen Metz, S. (2015). Why does the cognitive reflection test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition, 4*(3), 265-284. https://doi.org/10.1016/j.jarmac.2014.09.003

Barr, N., Pennycook, G., Stolz, J. A., & Fugelsang, J. A. (2015). Reasoned connections: A dual-process perspective on creative thought. *Thinking & Reasoning*, *21*(1), 61-75.

Brañas-Garza, P., García-Muñoz, T., & González, R. H. (2012). Cognitive effort in the Beauty Contest Game. *Journal of Economic Behavior and Organization*, *83*(2), 254–260.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*(2), 197-253. https://doi.org/10.1037/0033-2909.119.2.197

Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition*, *42*(3), 434-447.

Cho, S., Moody, T. D., Fernandino, L., Mumford, J. A., Poldrack, R. A., Cannon, T. D., Knowlton, B. J., & Holyoak, K. J. (2010). Common and dissociable prefrontal loci associated with component mechanisms of analogical reasoning. *Cerebral Cortex*, *20*, 524-533.

Cushen, P. J., & Wiley, J. (2018). Both attentional control and the ability to make remote associations aid spontaneous analogical transfer. *Memory & cognition*, *46*(8), 1398-1412.

Day, E. A., Espejo, J., Kowollik, V., Boatman, P. R., & McEntire, L. E. (2007). Modeling the links between need for cognition and the acquisition of a complex skill. *Personality and Individual Differences*, *42*(2), 201–212.

Don, H. J., Goldwater, M. B., Otto, A. R., & Livesey, E. J. (2016). Rule abstraction, model-based choice, and cognitive reflection. *Psychonomic Bulletin & Review*, *23*(5), 1615–1623.

Doumas, L. A. A., Morrison, R. G., & Richland, L. E. (2018). Individual differences in relational learning and analogical reasoning: A computational model of longitudinal change. *Frontiers in Psychology*, *9* (1235). https://doi.org/10.3389/fpsyg.2018.01235

Duncker, K. (1945). On problem solving. *Psychological Monographs*, Whole No. 58.

Fisher, K. J., Borchert, K., & Bassok, M. (2011). Following the standard form: Effects of equation format on algebraic modeling. *Memory & Cognition*, *39*(3), 502-515.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & cognition*, *43*(2), 226-236.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*(3), 306–355.

Goel, A. K. (1997). Design, analogy, and creativity. *IEEE Expert*, *12*(3), 62-70.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology, 23*(2), 222-262. https://doi.org/10.1016/0010-0285(91)90010-L

Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin*, *142*(7), 729–757.

Griffin, T. D., Wiley, J., Britt, M. A., & Salas, C. R. (2012). The role of CLEAR thinking in learning science from multiple-document inquiry tasks. *International Electronic Journal of Elementary Education*, *5*(1), 79-79.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*(6), 803-831-864.

Hanaki, N., Jacquemet, N., Luchini, S., & Zylbersztejn, A. (2016). Fluid intelligence and cognitive reflection in a strategic environment: Evidence from dominance-solvable games. *Frontiers in Psychology*, *7*(AUG), 1–12.

Hansell, N. K., Halford, G. S., Andrews, G., Shum, D. H., Harris, S. E., Davies, G., ... & Medland, S. E. (2015). Genetic basis of a cognitive complexity metric. *PloS one*, *10*(4), e0123886.

Haran, U., Ritov, I., & Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making, 8*(3), 188-201. Retrieved from https://search.proquest.com/docview/1420147630?accountid=14512

Hill, B. D., Foster, J. D., Elliott, E. M., Shelton, J. T., McCain, J., & Gouvier, W. D. (2013). Need for Cognition is related to higher general intelligence, fluid intelligence, and crystallized intelligence, but not working memory. *Journal of Research in Personality*, *47*(1), 22–25.

Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), Oxford handbook of thinking and reasoning (pp. 234–259). New York: Oxford University Press.

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

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26*(2), 107-129. https://doi.org/10.1016/0001-6918(67)90011-X

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*(3), 427–466.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, *110*(2), 220–264.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637-671.

Keane, M. T. (1997). What makes an analogy difficult? The effects of order and causal structure on analogical mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 946-967.

Kokis, J. V., Macpherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, *83*(1), 26-52.

Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., … Knowlton, B. J. (2008). Distraction during relational reasoning: The role of prefrontal cortex in interference control. *Neuropsychologia*, *46*(7), 2020–2032.

Kruglanski, A. W. (1989). Lay epistemics and human knowledge: Cognitive and motivational bases. Plenum Press, New York, NY. Retrieved from https://search.proquest.com/docview/617688570?accountid=14512

Kruglanski, A. W., & Mayseless, O. (1988). Contextual effects in hypothesis testing: The role of competing alternatives and epistemic motivations. *Social Cognition, 6*(1), 1-20. https://doi.org/10.1521/soco.1988.6.1.1

Kubricht, J. R., Lu, H., & Holyoak, K. J. (2017). Individual differences in spontaneous analogical transfer. *Memory & Cognition*, *45*(4), 576–588.

Lesage, E., Navarrete, G., & de Neys, W. (2013). Evolutionary modules and Bayesian facilitation: The role of general cognitive resources. *Thinking & Reasoning*, *19*(1), 27–53.

Lewis, A. B., & Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational psychology*, *79*(4), 363.

Little, J. L., & McDaniel, M. A. (2015). Individual differences in category learning: Memorization versus rule abstraction. *Memory & cognition*, *43*(2), 283-297.

Lu, H., Chen, D., & Holyoak, K. J. (2012). Bayesian analogy with relational transformations. *Psychological Review*, *119*, 617-648.

Lu, H., Wu, Y. N., & Holyoak, K. J. (2019). The emergence of analogy from relation learning. *Proceedings of the National Academy of Sciences, USA*, *116*, 4176-4181.

Macpherson, R., & Stanovich, K. E. (2007). Cognitive ability, thinking dispositions, and instructional set as predictors of critical thinking. *Learning and Individual Differences*, *17*(2), 115-127.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*(4), 431–467.

Martin, S. A., & Bassok, M. (2005). Effects of semantic cues on mathematical modeling: Evidence from word-problem solving and equation construction tasks. *Memory & Cognition*, *33*(3), 471–478.

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General*, *143*(2), 668–93.

Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., & Knowlton, B. J. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *Journal of Cognitive Neuroscience*, *16*(2), 260–271.

Oldrati, V., Patricelli, J., Colombo, B., & Antonietti, A. (2016). The role of dorsolateral prefrontal cortex in inhibition mechanism: A study on cognitive reflection test and similar tasks through neuromodulation. *Neuropsychologia, 91*, 499-508. https://doi.org/10.1016/j.neuropsychologia.2016.09.010

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014). Cognitive style and religiosity: The role of conflict detection. *Memory & Cognition*, *42*(1), 1-10.

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2016). Is the Cognitive Reflection Test a measure of both reflection and intuition? *Behavior Research Methods*, *48*(1), 341–348.

Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic bulletin & review*, *24*(6), 1774-1784.

Petty, R. E., & Jarvis, W. B. (1996). An individual differences perspective on assessing cognitive processes. In N. Schwarz, & S. Sudman (Eds.), Answering questions: Methodology for determining cognitive and communicative processes in survey research; answering questions: Methodology for determining cognitive and communicative processes in survey research (pp. 221-257, Chapter xi, 469 Pages) Jossey-Bass, San Francisco, CA. Retrieved from https://search.proquest.com/docview/618782242?accountid=14512

Raven, J. C. (1938). Progressive Matrices: A perceptual test of intelligence, 1938, individual form. London: Lewis.

Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning and instruction*, *16*(2), 92-103.

Roets, A., & Van Hiel, A. (2011). Item selection and validation of a brief, 15-item version of the need for closure scale. *Personality and Individual Differences, 50*(1), 90-94. https://doi.org/10.1016/j.paid.2010.09.004

Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. *Cognitive Psychology*, *8*(2), 165-190.

Stamenković, D., Ichien, N., & Holyoak, K. J. (2019). Metaphor comprehension: An individual-differences approach. *Journal of Memory and Language*, *105*, 108-118.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*(2), 342–357.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672–695.

Sweller, J. (2011). Cognitive load theory. In J. P. Mestre, & B. H. Ross (Eds.), The psychology of learning and motivation: Cognition in education (vol. 55); the psychology of learning and motivation: Cognition in education (vol. 55) (pp. 37-76, Chapter xiv, 313 Pages) Elsevier Academic Press, San Diego, CA. https://doi.org/10.1016/B978-0-12-387691-1.00002-8 Retrieved from https://search.proquest.com/docview/1021142093?accountid=14512

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision making*, *11*(1), 99-113.

Tohill, J. M., & Holyoak, K. J. (2000). The impact of anxiety on analogical reasoning. *Thinking and Reasoning*, *6*(1), 27–40.

Toplak, M., West, R., & Stanovich, K. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275–1289.

Van Hiel, A., & Mervielde, I. (2003). The need for closure and the spontaneous use of complex and simple cognitive structures. *The Journal of Social Psychology, 143*(5), 559-568. https://doi.org/10.1080/00224540309598463

Vendetti, M. S., Wu, A., & Holyoak, K. J. (2014). Far-out thinking: Generating solutions to distant analogies promotes relational thinking. *Psychological Science*, *25*(4), 928–933.

Viskontas, I. V., Morrison, R. G., Holyoak, K. J., Hummel, J. E., & Knowlton, B. J. (2004). Relational integration, inhibition, and analogical reasoning in older adults. *Psychology and Aging, 19*(4), 581-591. https://doi.org/10.1037/0882-7974.19.4.581

Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition*, *28*(7), 1205–1212.

West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology, 100*(4), 930-941. https://doi.org/10.1037/a0012842