Available online at www.sciencedirect.com
SCIENCE DIRECT®
Personality and Individual Differences 38 (2005) 1149–1161
PERSONALITY AND INDIVIDUAL DIFFERENCES
www.elsevier.com/locate/paid

Thinking about personal theories: individual differences in the coordination of theory and evidence

Walter C. Sá [a,*], Carol N. Kelley [b], Caroline Ho [b], Keith E. Stanovich [b]

[a] *Department of Psychology, Grand Valley State University, 1 Campus Drive, Allendale, Michigan 49401-9403, USA*
[b] *University of Toronto*

Received 10 February 2004; received in revised form 17 June 2004; accepted 9 July 2004

Abstract

In this study, we examined individual differences on an informal reasoning task that involved the coordination of theory and evidence. Using a structured interview paradigm adapted from the work of Kuhn (1991), a total of 96 participants were asked to generate evidence for two different self-generated theories. Results showed that individuals with high and low cognitive ability scores generated very similar types and patterns of evidence to the interview questions. The tendency to use the most sophisticated type of evidence in argument (covariation comparison) was not related to cognitive ability or thinking dispositions. However, the tendency to use an unsophisticated form of non-evidence (reiteration or elaboration of the original theory) was associated with both lower cognitive ability and lower actively open-minded thinking.

Keywords: Reasoning; Critical thinking; Cognitive styles; Thinking dispositions; Intelligence; Informal reasoning; Rationality; Decontextualized thinking

* Corresponding author. Tel.: +1 616 331 3705; fax: +1 616 331 2480.
E-mail address: saw@gvsu.edu (W.C. Sá).

0191-8869/$ - see front matter
doi:10.1016/j.paid.2004.07.012

1. Introduction

Many theorists view the ability to calibrate beliefs appropriately to evidence as one of the most fundamental critical thinking skills (e.g., Baron, 2000; Kuhn, 1991). Other investigators have emphasized the importance of decontextualized modes of thinking when attempting to calibrate the strength of belief to the nature of the evidence available (Baron, 1995, 2000; Perkins, 1995; Perkins, Jay, & Tishman, 1993; Stanovich, 1999, 2004). Decontextualization skills enable reasoning processes to operate independent of interfering context. For example, researchers have studied thinking skills such as the ability to overcome belief bias (accepting or rejecting an argument's conclusion based on its believability rather than actual logical validity), the ability to adopt stances other than one's own, and the ability to avoid myside bias (generating more arguments in favor of a one's own position than arguments opposing that position) when evaluating evidence (Baron, 1995; Perkins, Farady, & Bushey, 1991; Sá, West, & Stanovich, 1999; Toplak & Stanovich, 2003).

Researchers have also begun to examine individual differences that are associated with these important critical thinking skills. For example, it is important to establish the extent to which these critical thinking skills are merely proxies for cognitive ability or whether they reflect thinking dispositions that are dissociable from intelligence. In a series of studies, Stanovich and colleagues (Sá & Stanovich, 2001; Sá et al., 1999; Stanovich & West, 1997) found that critical aspects of decontextualization—particularly the ability to avoid belief bias—are indeed predicted by measures of cognitive ability. Nevertheless, in these same studies, it was found that thinking dispositions related to epistemic regulation (e.g., belief-related dispositions such as dogmatism, belief identification, and categorical thinking) predicted the tendency to avoid belief bias even after cognitive ability was partialled out.

In several studies by Klaczynski and colleagues (Klaczynski, 1997; Klaczynski & Gordon, 1996; Klaczynski & Robinson, 2000; Klaczynski, Gordan, & Fauth, 1997), the data patterns were even more clear-cut. These investigators found that various forms of decontextualized reasoning (in particular, aspects of myside bias) were predicted by thinking dispositions related to epistemic regulation. However, they also found that several aspects of decontextualized reasoning were independent of cognitive ability (Klaczynski, 1997; Klaczynski & Gordon, 1996; Klaczynski & Robinson, 2000; Klaczynski et al., 1997).

Thus, there is convergence in the literature on the link between unbiased reasoning and thinking dispositions, but there is disagreement about whether cognitive ability is linked to decontextualized reasoning. One important difference between the two research programs is that the informal reasoning tasks utilized by the Klaczynski group are not accompanied by specific instructions to decontextualize. In contrast, in the Sá et al. (1999) study; (see also Stanovich & West, 1997), participants are given directive instructions to decontextualize—to set aside their prior beliefs and opinions and to engage in bias-free reasoning.

In the present study, we examine these individual difference relationships in a paradigm like that used by the Klaczynski group—one that does not utilize specific instructions to decontextualize. The task used here is a variant of the structured interview task introduced by Kuhn (1991) that focuses on the ability to coordinate theory and evidence. In two separate interviews, participants were asked to generate a causal theory. In the first interview, participants are asked for a causal theory of why students fail in school. In the second interview, participants are asked for a

causal theory of why prisoners return to crime after they are released from jail. The responses generated to support their personal theories provided us with an opportunity to examine the types of evidence and counter evidence people generate in such informal reasoning situations. Importantly, in the Kuhnian interview, the causal theories generated by the participants are not the focus of the investigation. Rather, it is the subsequent evidence participants generate to support the theories they articulate that is the focus.

A crucial point about these Kuhnian interviews is that they yield categories of responses that vary in sophistication. We were particularly interested in the generation of a popular form of "nonevidence" reported by Kuhn (1991). Kuhn noted that a very common response to a request for evidence in support of a theory was a mere reiteration or elaboration of the original theory itself. Kuhn refers to this as pseudoevidence since this sort of response cannot be differentiated from the theory it purports to support. The key to the Kuhnian interview is to understand that theory and evidence are conceptually decoupled. Importantly for us, a participant who gives many reiteration/elaboration responses is indicating that they fail to understand this decoupling. In contrast, a clear example of an evidence type which indicates that participants do understand the decoupled nature of theory and evidence is provided when participants implicate some sort of covariation comparison (implicating the importance of finding some covariation between the outcome at issue and the suggested cause) as a source of evidence. Kuhn (1991) labels this type of response as genuine evidence because it "(a) is differentiated from the theory... and (b) bears on the theory's correctness" (Kuhn, 1992, p. 159).

Thus, in the Kuhnian interview, there are two critical metrics for the quality of reasoning. The first, the number of reiteration/elaboration responses, is an indicator of unsophisticated reasoning or, at best, a shallow understanding of what constitutes evidence. The second, responses making reference to covariation comparison, represents sophisticated reasoning about what type of evidence is needed to support a claim. In the present study, we examine whether these indices of critical thinking are associated with measures of cognitive ability and thinking dispositions. The pattern of correlations obtained speak importantly to the issue of whether the critical thinking skills discussed in the reasoning literature actually reflect cognitive styles that are different from the cognitive capacities measured on intelligence tests. Unlike the majority of prior studies of reasoning which employ university students, our investigation examined participants who varied widely in age and were returning to school for their high school credential. Although still restricted in range compared to the general population, this group will display, if anything, more variability than a student sample.

2. Method

2.1. Participants

The participants were 96 mature students (47 males and 49 females) who had returned to school to complete their high school education. Their mean age was 32.0 years (SD = 9.4) and their ages ranged from 19 to 57 years old. Participants were paid $20 for their involvement. One participant did not complete the thinking dispositions questionnaire, and one further participant was removed as a multivariate outlier (see Tabachnick & Fidell, 1983).

2.2. Theory/evidence interviews

The theory/evidence interview reported here is heavily based on a research task/interview developed by Kuhn (1991). Participants were administered a prestructured interview twice. Both interviews had identical protocols. Each interview centered on a different issue: ''what causes children to fail in school?'' (referred to hereafter as the school issue), and ''what causes prisoners to return to crime after they are released?'' (referred to hereafter as the crime issue). The protocol of these interviews will be described by using the school issue as an example.

Participants were first asked the causal theory question ''what causes students to fail in school?'' Participants who gave more than one cause were then asked ''which of these, that you just said, would you say is the major cause of children failing in school?'' The interview from this point is focused on a single causal theory about school failure, and all subsequent questions center on this focal theory generated by the participant.

Two separate lines of questioning then follow. Both lines of questioning, consisted of a core set of questions that were consistently asked of participants, and optional questions that were asked at the experimenter's discretion. We report here the core questions only, and our subsequent analyses are restricted to the responses participants made in reference to these core questions.

The first line of questioning consisted of two direct requests for evidence: ''How do you know that this is the cause?'' followed by ''If you were trying to convince someone else that your view, [focal theory repeated here], is right, what evidence would you give to try to show this?'' These two questions are then followed by a second line of questioning which requires some element of considering a theory that conflicts with the participant's own focal theory: ''Now suppose that someone disagreed with your view that this is the cause, [focal theory repeated], what might they say to show that you were wrong?''; ''What evidence might this person give to try to show that you were wrong?''; ''Could someone prove that you were wrong?''; ''Suppose that someone disagreed with you and said that the [an alternative theory is suggested here by the experimenter] was the cause, okay? What could you say to show that this person was wrong?''; ''Would you be able to prove this person wrong?'' These questions were asked in the order reported above.

The responses participants made to each of these questions were coded using a prestructured coding scheme (see Table 1 for summary of these codes). Seven code classifications were relevant to all questions. Reiteration or Elaboration is assigned to responses in which participants simply reiterate their theory or elaborate upon their theory in response to a question (e.g., when asked to provide evidence for why ''not paying attention'' was the major cause of school failure for children, participant 11 responded: ''Because if you don't pay attention, you don't know what's going on''—i.e., a response that is a mere elaboration upon the previously stated theory). Covariation Comparison is assigned to responses in which participants implicate some sort of covariation comparison as a source for evidence (e.g, asked for some evidence for why ''parental influence'' was the major cause of school failure for children, participant 85 responded: ''I would perhaps do a survey, go to a high school and see, and ask students just that one question if their parents are instilling education... you know good education in their minds, and how they're doing, what their grade point average is''). The Personal Experience classification is assigned to responses in which the participant implicates the importance of personal experience (e.g., in response to the question ''How do you know that this is the cause?'' participant 18 stated: ''Cuz it was my...I'm relating to past experiences, when I was younger''). The External Source code is assigned to re-

Table 1
Evidence type codes used to classify participants' responses

Codes Relevant to All Questions	
Reiteration or Elaboration:	Participant simply reiterates a previous causal theory, or elaborates on their previously stated causal theory
Covariation Comparison:	Participant implicates or calls for some specified covariation comparison between the alleged causal antecedent and the outcome at issue
Personal Experience:	Participant implicates the importance of personal experience by either relating (an) experience(s), or stating that they have no personal experience on which to draw a conclusion
External Source:	Participant implicates the importance of some form of external source of information by either stating that the relevant people should be asked about what the cause is, or a nonspecific call for statistics, experiments, surveys, expert opinion
Opinion:	Participant does not produce evidence because "it's only my opinion" or "just my view"
Rebuttal:	A rebuttal of an alternative theory that conflicts with the participant's original theory Alternative theory may have been suggested by the experimenter, or generated by participant in response to a prompt by experimenter
Don't Know:	Participant states that they don't know; or makes no response; or refuses or balks
Codes Relevant to Following Two Questions Only	
"Could someone prove that you were wrong?"	
"Would you be able to prove this person wrong?"	
Unequivocal yes:	Participant responds with an unequivocal yes
Unequivocal no:	Participant responds with an unequivocal no
Equivocal yes:	Participant responds with an equivocal yes
Equivocal no:	Participant responds with an equivocal no

sponses in which the participant implicates the importance of some form of external source of information (e.g., for the crime issue, participant 33 suggested asking the relevant people: "... you can ask criminals in jail. Ask them what they're in jail for, and why did they do that"; on the same issue and question, participant 64 made a nonspecific call for statistics, surveys etc.: "I guess I'd have to go and find some factual evidence, you know, surveys that's already been taken ... maybe go to the library or um maybe there's places you can call where there's family, you know, family centers I think where you can call up and get some statistics, figures and stuff"). The Opinion classification is assigned to responses in which a participant does not produce evidence and states that her/his view is "only an opinion" (e.g., participant 1 stated "I'm just aah, assuming and aah, it's an opinion..."). The Rebuttal code was assigned when a participant attempted to contradict a theory that was in opposition to the originally stated theory (e.g., participant 90 suggested that "students skipping school" was the cause of school failure and responded to an alternative "poor teachers" theory with the following "... but I don't think it would stand up because it is, in a fact a teacher can be there and can be mean and can do whatever it wants or she or he wants, but in the end she's still there and at the end of the class and if you're not you're the one skipping not her or him"). Finally, the Don't Know classification was assigned to responses where the participant makes no response, or simply states s/he does not know (e.g., participant 19 responded to a question with "How do I know?...I don't know how to answer that", participant

58 responded to a question with "I can't think of evidence"). Two of our questions also had unique response codes displayed on the bottom of Table 1.

The interviews were recorded on audio tape and were later transcribed on to paper. Two separate coders were trained on the coding scheme and independently coded each of the transcripts. The coders were asked to indicate which of the codes (viz., evidence types) applied to participants' responses. More than one code could apply to a single response. The coders converged on 85% of these responses. Discrepancies between the two coders were rectified by having a third coder review the transcripts and determine the final code.

2.3. Cognitive ability measure

We assessed cognitive ability using both a verbal and nonverbal task. The former was a commonly used measure of crystallized intelligence and the latter a classic measure of fluid intelligence (Horn & Noll, 1997). We combined performance on these tasks in order to obtain a comprehensive overall measure of cognitive ability. Participants were administered 20 problems from Raven's Standard Progressive Matrices (Raven, 1960). The participants were given 15 min to complete the 20 items on the test. These 20 items consisted of seven items taken from Set C, seven items from Set D, and six items taken from Set E. A brief vocabulary measure was also administered to participants (vocabulary is the strongest specific correlate of general intelligence, see Matarazzo, 1972). The verbal measure employed the checklist-with-foils format that has been shown to be a reliable and valid way of assessing individual differences in vocabulary knowledge (Anderson & Freebody, 1983; Stanovich, West, & Harrison, 1995). The stimuli for the task were 40 words (e.g., absolution, irksome, purview) and 20 pronounceable nonwords (e.g., disler, potomite, seblement) taken largely from the stimulus list of Zimmerman, Broder, Shaughnessy, and Underwood (1977). The words and nonwords were intermixed via alphabetization. The participants were told that some of the letter strings were actual words and that others were not and that their task was to read through the list of items and to put a check mark next to those that they knew were words. Scoring on the task was determined by taking the proportion of the target items that were checked and subtracting the proportion of foils checked.

One cognitive ability variable reported in this study was created by standardizing the raw scores on the Raven's Matrices and vocabulary checklist measure and then adding the two standard scores together (there was a significant correlation between Raven's Matrices and the vocabulary checklist $r = .35$, $p < .001$). This resulted in a cognitive ability variable with continuous scores. In addition to using this continuous variable, our analyses include the use of a dichotomous cognitive ability variable that was formed by sorting the above continuous variable on magnitude of score and then performing a median-split resulting in a low cognitive ability group ($N = 47$) and a high cognitive ability group ($N = 48$).

2.4. Thinking disposition questionnaire

Two thinking disposition scales were formed: a 36-item Actively Open-minded Thinking scale and a 9-item Superstitious Thinking scale. The response format for each item was agree strongly (scored as 6), agree moderately (5), agree slightly (4), disagree slightly (3), disagree moderately (2),

disagree strongly (1). Items from these two scales were randomly intermixed with each other and with other scales not part of the present investigation.

2.5. Actively open-minded thinking scale

This scale was formed by using a variety of items from the following measures:

Eight items from the Openness-Values facet of the Revised NEO Personality Inventory (Costa & McCrae, 1992) were administered (e.g., "I believe that laws and social policies should change to reflect the needs of a changing world"); 9 items (reverse scored) measuring dogmatism (e.g., "No one can talk me out of something I know is right"; Paulhus & Reid, 1991; Robinson, Shaver, & Wrightsman, 1991; Troldahl & Powell, 1965); 10 items measuring flexible thinking (e.g., "People should always take into consideration evidence that goes against their beliefs"; see Stanovich & West, 1997); and 9 items (reverse scored) measuring what we have termed belief identification ("It makes me happy and proud when someone famous holds the same beliefs that I do"; see Sá et al., 1999). See Stanovich and West (1997) and Sá et al. (1999) for further examples of items on the Actively Open-minded Thinking scale. The score on the scale is obtained by summing the responses to the 36 items. The split-half reliability of the Actively Open-minded Thinking scale (Spearman–Brown corrected) was .78. Cronbach's alpha was .79. High scores on the Actively Open-minded Thinking scale indicate openness to belief change and cognitive flexibility; whereas low scores indicate cognitive rigidity and resistance to belief change.

2.6. Superstitious thinking scale

We also included a Superstitious Thinking scale as a further thinking disposition index germane to the appropriateness of belief formation in light of relevant evidence. This scale was formed by using a variety of items from the following measures:

Two items were concerned with belief in astrology (e.g., "It is advisable to consult your horoscope daily"); and were adapted from the paranormal belief scale validated by Jones, Russell, and Nickel (1977); Four items reflecting conceptions of luck were taken from previous work by Stanovich and West (1998, e.g., "The number 13 is unlucky") and were similar to items on the superstition subscale of a paranormal beliefs questionnaire developed by Tobacyk and Milford (1983); Three items from the Superstitious Thinking subscale of Epstein and Meier's (1989) Constructive Thinking Inventory (e.g., "I have found that talking about successes that I am looking forward to can keep them from happening"). The Superstitious Thinking scale was formed by summing the raw scores from these items. The split-half reliability of the Superstitious Thinking scale (Spearman–Brown corrected) was .78. Cronbach's alpha was .72. Higher scores on this scale indicate increased levels of superstitious thinking.

2.7. Procedure

The tasks (along with some other tasks and measures not part of the present investigation) were administered over two testing sessions on two separate days by the same experimenter. Participants were tested in small groups on the first day's test session. During this first session participants completed a demographics form, Raven's Standard Progressive Matrices, and the

thinking disposition questionnaire. On the second testing session, participants were individually administered the theory/evidence interviews.

3. Results

The first column of Table 2 presents the mean number of each evidence type generated by our participants across both the school and crime issues and all questions. As discussed in the Introduction, two of these categories are of particular interest. The category of Reiteration or Elaboration is an indicator of unsophisticated reasoning or, at best, a shallow understanding of what constitutes evidence. In contrast, Covariation Comparison represents sophisticated reasoning about what type of evidence is needed to support a theory. Although the level of sophistication for five of the seven evidence codes is ambiguous, it is clear that as evidence, Reiteration or Elaboration is of poor quality and Covariation Comparison is sophisticated. Thus, it is important to note that there are indications in Table 2 that the generation of Reiteration or Elaboration—viz., poor quality evidence—is more frequent than the generation of Covariation Comparison—viz., a sophisticated evidence type. Seventy out of 95 participants provided more Reiteration or Elaboration than Covariation Comparison. A within-subjects t-test indicated that across all participants significantly more Reiteration or Elaboration responses were made than Covariation Comparison responses ($t(94) = 3.95$, $p < .001$). A repeated measures ANOVA was also performed on the frequency of all seven evidence types. The omnibus repeated measures ANOVA indicated significant differences, $F(6, 94) = 7.57$, $p < .001$. A series of post-hoc Fisher PLSD tests indicated that the frequency of Covariation Comparison—what Kuhn has called genuine evidence—was significantly lower than all evidence codes with the exception of the Don't Know code (where it was statistically no different in frequency).

We also tested the independence of evidence type generation across the two issues. A median-split on the total number of Reiteration or Elaboration responses (High Reiteration or Elaboration versus Low Reiteration or Elaboration) made during the school issue was compared with the analogous median-split of the same evidence type generated in the crime issue. Results indicate

Table 2
Mean scores of all participants and participants with highest and lowest scores on the cognitive ability measure (standard deviations in parentheses)

Evidence type	All ($N = 95$)	Low CA ($N = 47$)	High CA ($N = 48$)	t-Value
Reiteration or Elaboration	2.64 (1.66)	2.92 (1.76)	2.38 (1.53)	1.76 (ns)
Covariation Comparison	1.59 (1.65)	1.47 (1.60)	1.71 (1.71)	−.71 (ns)
Personal Experience	2.60 (1.97)	2.62 (2.10)	2.58 (1.89)	.08 (ns)
External Source	2.67 (2.34)	2.51 (2.37)	2.83 (2.33)	−.67 (ns)
Opinion	2.39 (1.12)	2.38 (1.23)	2.40 (1.02)	−.06 (ns)
Rebuttal	2.27 (1.55)	2.34 (1.46)	2.21 (1.65)	.41 (ns)
Don't Know	1.45 (1.62)	1.30 (1.41)	1.60 (1.80)	−.92 (ns)

Note: Low CA = low cognitive ability group resulting from median-split on the continuous cognitive ability score, High CA = high cognitive ability group resulting from median-split on the continuous cognitive ability score; t values correspond to t tests performed on each evidence type between the Low CA and High CA groups.

that the generation of this evidence type across the two issues was not independent $\chi^2(1, N = 95) = 5.32$, $p < .05$. Participants who generated higher frequencies of Reiteration or Elaboration in the school issue were also more likely to generate higher frequencies of the same evidence type in the crime issue. The same sort of analysis was performed for each of the remaining evidence types. In all cases the number of instances in which we observed the corresponding evidence types across the two issues was higher than would be expected under an assumption of independence (Covariation Comparison, $\chi^2(1, N = 95) = 6.91$, $p < .01$; Personal Experience, $\chi^2(1, N = 95) = 9.84$, $p < .01$; External Source, $\chi^2(1, N = 95) = 13.55$, $p < .001$; Opinion, χ^2 (1, $N = 95) = 10.47$, $p < .01$; Rebuttal, $\chi^2(1, N = 95) = 4.10$, $p < .05$; and Don't Know, $\chi^2(1, N = 95) = 9.62$, $p < .01$).

The next two columns of Table 2 indicate the mean evidence type scores for individuals of low and high cognitive ability. A series of *t*-tests were used to compare the frequency of each evidence type generated by the high and low cognitive ability groups. The low cognitive ability group was statistically no different than the high cognitive ability group on any of the evidence types. The response patterns of these two groups are quite similar. For example, both groups gave more Reiteration or Elaboration responses than Covariation Comparison responses and both groups made Covariation Comparison responses less frequently than all other responses with the exception of the Don't Know response.

A 2×2 mixed ANOVA was performed with cognitive ability group (low vs. high) entered as a between group factor and Reiteration or Elaboration /Covariation Comparison entered as a within group factor. Of interest here was an examination of whether there was any interaction between Cognitive Ability and the generation of Reiteration or Elaboration/Covariation Comparison. The interaction effect was not statistically significant, $F(1, 93) = 1.52$, $p > .05$.

Of course, the discrete dichotomy of the median-split loses some information. Fully continuous correlational data are presented in Table 3. That Table focuses on the Reiteration or Elaboration and Covariation Comparison responses (i.e., the evidence types that are nonambiguous in terms of sophistication); cognitive ability, the Actively Open-minded Thinking scale and Superstitious Thinking. The category representing the most sophisticated type of reasoning, Covariation Comparison, shows no relationship with any of the individual difference variables. Thus, regarding

Table 3
Intercorrelations among selected evidence types, cognitive ability, actively open-minded thinking and superstitious thinking

Variable	1	2	3	4
1. Reiteration or Elaboration				
2. Covariation Comparison	−.242*			
3. Cognitive Ability	−.253*	.000		
4. Actively Open-minded Thinking	−.318**	.043	.652***	
5. Superstitious Thinking	.206*	−.053	−.403***	−.511***

Note: $N = 94$ for comparisons involving either the Actively Open-minded Thinking scale or the Superstitious Thinking scale; $N = 95$ for remaining comparisons.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

cognitive ability, the results of the median split analysis and the correlational analysis converged. Participants high in cognitive ability do not display a tendency to give increased levels of the more sophisticated Covariation Comparison responses.

However, cognitive ability did display a significant negative correlation with Reiteration or Elaboration, indicating that those higher in cognitive ability tend to give fewer Reiteration or Elaboration responses. The Actively Open-minded Thinking score attained an even stronger negative correlation with Reiteration or Elaboration (see Table 3). In fact, the Actively Open-minded Thinking Score predicted variation in Reiteration or Elaboration even after cognitive ability was partialled out (partial $r = -.21$, standardized Beta = $-.267$, unique variance explained = .041, $p < .05$). The opposite was not true. Cognitive ability did not predict variation in Reiteration or Elaboration after partialling out the variation accounted for by Actively Open-minded Thinking (partial $r = -.06$, standardized Beta = $-.079$, unique variance explained = .004, $p > .05$). The Superstitious Thinking scale was also found to correlate positively with Reiteration or Elaboration. No associations between the remaining five evidence types were found with our individual difference variables.

We also explored relationships between our focal variables and various demographic variables (e.g., education level, reading habits, gender, age etc.). Age proved to be the only demographic variable displaying any relationship with our focal variables. Age was negatively correlated with the generation of the Covariation Comparison evidence type, $r = -.30$, $p < .01$, (cf., Klaczynski & Robinson's (2000) discussion of the tendency to engage in heuristic processing increasing with age). Although not significantly related to our cognitive ability measure (i.e., a verbal and nonverbal composite score), age was negatively correlated to the raw score on Raven's Standard Progressive Matrices ($r = -.35$, $p < .001$) and showed a positive trend with the raw score on the vocabulary measure ($r = .20$, $p < .10$). A lack of variability in education level—all our participants had not completed high school—limited the value of any comparisons using this variable and may account for it not being related to any of our focal variables.

Finally, two of the questions in the interview had an additional set of unique response classifications. Responses to the questions "Could someone prove that you were wrong?" and "Would you be able to prove this person wrong?" were classified as either being an Unequivocal Yes, Unequivocal No, Equivocal Yes, or Equivocal No (see Table 1). Unequivocal No responders to the question "Could someone prove that you were wrong?" were related to our individual difference variables in an interesting way. Participants who responded with an Unequivocal No for the above question in at least one of the two issues ($N = 22$) were compared to those that did not provide an Unequivocal No response in either issue ($N = 72$–73). The Unequivocal No group scored significantly lower on the cognitive ability measure, $t(93) = 2.25$, $p < .05$, scored significantly lower on the Actively Open-minded Thinking scale, $t(92) = 2.26$, $p < .05$, and significantly higher on the Superstitious Thinking scale, $t(92) = -2.90$, $p < .01$. There were no other patterns between these unique codes and our individual difference measures.

4. General discussion

The present study found that on an informal reasoning task requiring the generation of evidence and counter evidence, the responses of participants of high and low cognitive ability dis-

played many similarities. Both groups were significantly more likely to reiterate or elaborate on a previously stated theory when asked for evidence supporting their theory than they were to provide a form of genuine evidence. Thus, our results suggest a similarity of informal reasoning styles across variation in cognitive ability (see Perkins, 1985, for a related discussion on the similarity of informal reasoning styles across a spectrum of education level).

Of course, our conclusions must be tempered by the fact that we did not sample the entire range of cognitive ability levels. However, there was considerable variability in our sample (the high ability participants solved an average of 7.4 more Raven problems than the low ability participants). In a study using a different selection of Raven items with university students (Sá et al., 1999), the standard deviation (3.33) was lower than that in this study (4.60). Additionally, in our continuous analyses, we were able to find some differences between high and low ability participants. Cognitive ability was negatively correlated with the generation of reiterations or elaborations. In addition, the frequency of reiteration/elaboration was negatively correlated with a measure of actively open-minded thinking. Whereas actively open-minded thinking was found to predict unique variance in reiteration or elaboration after statistically controlling for cognitive ability, cognitive ability did not account for unique variance in the reiteration or elaboration response after controlling for actively open-minded thinking.

Furthermore, reiteration or elaboration was also correlated positively with a measure of superstitious thinking. Thus an important skill in decontextualization (i.e., decoupling evidence from theory—or at least the tendency to avoid a response that conflates these two) was found to be associated with both cognitive ability and two thinking disposition scales. The relationship with actively open-minded thinking remained significant despite statistically controlling for cognitive ability. Thus, the tendency to avoid the unsophisticated reiteration/elaboration responses appears to reflect a cognitive style that is separate from the cognitive capacities measured on intelligence tests. In contrast, the frequency of covariation comparison—which can be viewed as a sophisticated form of evidence—did not correlate with our individual difference variables.

The potency of thinking dispositions over cognitive ability as a predictor of a decontextualized thinking skill in this study is unlike that found in a previous series of studies by Stanovich and colleagues (e.g., Sá et al., 1999; Stanovich & West, 1997). There, cognitive ability remained a significant predictor after thinking dispositions were partialled. This difference may have resulted because the informal reasoning task used in the present investigation—unlike those used in the previous series cited above—is not accompanied with specific instructions to engage in a decontextualized mode of thinking. Cognitive ability may be increasingly important when the participant is supplied with instructions about what needs to be done in order to reason effectively. In the absence of these external instructions, thinking dispositions become particularly potent because they will determine how an individual's cognitive ability is to be allocated when the task demands are ambiguous.

Finally, we observed that the willingness to concede that one might be proven wrong was associated with our individual difference variables. Kuhn (1991, 1992) had previously reported that some participants refused to acknowledge the possibility that their theory may be proven wrong. A prerequisite to having beliefs adequately track their evidential basis would seem to be a willingness to acknowledge that beliefs are tentative in nature. Nevertheless, our results indicate that lack of cognitive ability does not prevent individuals from accomplishing some degree of decoupling from belief. The profile of argument types was similar for individuals of high and low cognitive

ability (see Table 2). Additionally, as indicated in Table 3, the tendency to give evidence that is independent of theory (covariation comparison) displayed no association with cognitive ability. Although the tendency to provide nondiagnostic evidence (reiteration/elaboration) was correlated with cognitive ability, this correlation did not survive statistical control for an important thinking disposition—actively open-minded thinking. The latter thinking disposition was more strongly related to the tendency to give nondiagnostic evidence than was intelligence. Critical thinking in this informal reasoning situation seems to be best conceptualized as a dispositional construct than as a cognitive capacity (Baron, 1995; Dole & Sinatra, 1998; Klaczynski & Robinson, 2000; Kuhn, 1991; Perkins et al., 1993; Stanovich, 1999, 2004).

Acknowledgements

This research was supported by grants from Social Sciences and Humanities Research Council of Canada and the Canada Research Chairs program to Keith E. Stanovich.

We thank Carole Leckner, Robin Sidhu, Christine Tardif, Maggie Toplak and Simon Williams for transcribing the audio taped interviews. We also thank Christine Tardif and Simon Williams for coding the interviews. An anonymous reviewer is thanked for his/her comments on this manuscript.

References

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Huston (Ed.), *Advances in reading/language research* (Vol. 2, pp. 231–256). Greenwich, CT: JAI Press.

Baron, J. (1995). Myside bias in thinking about abortion. *Thinking and Reasoning, 1*, 221–235.

Baron, J. (2000). *Thinking and deciding* (3rd ed.). Cambridge, MA: Cambridge University Press.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory*. Odessa, FL: Psychological Assessment Resources.

Dole, J. A., & Sinatra, G. M. (1998). Reconceptualizing change in the cognitive construction of knowledge. *Educational Psychologist, 33*, 109–128.

Epstein, S., & Meier, P. (1989). Constructive thinking: a broad coping variable with specific components. *Journal of Personality and Social Psychology, 57*, 332–350.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf–Gc theory. In D. Flanagan, J. Genshaft, & P. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (pp. 53–91). New York: Guilford Press.

Jones, W., Russell, D., & Nickel, T. (1977). Belief in the paranormal scale: An objective instrument to measure belief in magical phenomena and causes. *JSAS Catalog of Selected Documents in Psychology, 7* (100 (Ms. No. 1577)).

Klaczynski, P. A. (1997). Bias in adolescents' everyday reasoning and its relationship with intellectual ability, personal theories, and self-serving motivation. *Developmental Psychology, 33*, 273–283.

Klaczynski, P. A., & Gordon, D. H. (1996). Self-serving influences on adolescents' evaluations of belief-relevant evidence. *Journal of Experimental Child Psychology, 62*, 317–339.

Klaczynski, P. A., Gordan, D. H., & Fauth, J. (1997). Goal-oriented critical reasoning and individual differences in critical reasoning biases. *Journal of Educational Psychology, 89*, 470–485.

Klaczynski, P. A., & Robinson, B. (2000). Personal theories, intellectual ability, and epistemological beliefs: adult age differences in everyday reasoning tasks. *Psychology and Aging, 15*, 400–416.

Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review, 62*, 155–178.
Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). Baltimore: The Williams & Wilkins Co.
Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*, 307–317.
Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology, 77*, 562–571.
Perkins, D. N. (1995). *Outsmarting IQ: the emerging science of learnable intelligence*. New York: Free Press.
Perkins, D. N., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. Voss, D. Perkins, & J. Segal (Eds.), *Informal reasoning and education* (pp. 83–105). Hillsdale, NJ: Erlbaum.
Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: a dispositional theory of thinking. *Merrill–Palmer Quarterly, 39*, 1–21.
Raven, J. C. (1960). *Standard progressive matrices (Set B)*. London: H.K. Lewis & Co.
Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). *Measures of personality and social psychological attitudes* (Vol. 1). San Diego, CA: Academic Press, pp. 560–564.
Sá, W. C., & Stanovich, K. E. (2001). The domain specificity and generality of mental contamination: accuracy and projection in judgments of mental content. *British Journal of Psychology, 92*, 281–302.
Sá, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 91*(3), 497–510.
Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates.
Stanovich, K. E. (2004). *The robot's rebellion: finding meaning in the age of Darwin*. Chicago: University of Chicago Press.
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*, 342–357.
Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*, 161–188.
Stanovich, K. E., West, R. F., & Harrison, M. (1995). Knowledge growth and maintenance across the life span: the role of print exposure. *Developmental Psychology, 31*, 811–826.
Tabachnick, B. G., & Fidell, L. S. (1983). *Using multivariate statistics*. New York: Harper & Row.
Tobacyk, J., & Milford, G. (1983). Belief in paranormal phenomena. *Journal of Personality and Social Psychology, 44*, 1029–1037.
Toplak, M. E., & Stanovich, K. E. (2003). Associations between myside bias on an informal reasoning task and amount of post-secondary education. *Applied Cognitive Psychology, 17*, 851–860.
Troldahl, V., & Powell, F. (1965). A short-form dogmatism scale for use in field studies. *Social Forces, 44*, 211–215.
Zimmerman, J., Broder, P. K., Shaughnessy, J. J., & Underwood, B. J. (1977). A recognition test of vocabulary using signal-detection measures, and some correlates of word and nonword recognition. *Intelligence, 1*, 5–31.