Title
Working Memory and Causal Reasoning under Ambiguity

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN
1069-7977

Authors
Shou, Yiyun
Smithson, Michael

Publication Date
2014

Peer reviewed

Working Memory and Causal Reasoning under Ambiguity

Yiyun Shou (yiyun.shou@anu.edu.au)
Michael Smithson (michael.smithson@anu.edu.au)
Research School of Psychology, The Australian National University
Canberra, ACT, Australia

Abstract

Causal reasoning involves evaluation and integration of the observed evidence, the quality of which is influenced by the external factors such as uncertainty and the internal factors such as one's cognitive ability. The current experimental study investigated the relationship between working memory (WM), causal reasoning and impacts of ambiguous observations. Results revealed that WM assessed by the *n*-back task was associated with subjects' causal reasoning under unambiguous condition. The higher *n*-back scores were associated with lower variability in causal ratings. On the other hand, WM assessed by the operational span task was associated with subjects' reaction to the ambiguous evidence. Subjects with higher span had greater individual difference in their reactions to the ambiguous evidence than those with lower WM capacity.

Keywords: Causal reasoning; Working memory; Ambiguity; Uncertainty

Introduction

The formation of causal beliefs requires the integration of the observed causal cues and making inferences about the causal relationships between two variables. Cognitive resources such as attention and memory are essential for the underlying reasoning process that involves holding and evaluating the observed evidence (Buehner, Krumm, & Pick, 2005). Working memory (WM) refers to the ability to retain and actively manipulate information simultaneously, and working memory capacity (WMC) is currently a key indicator of the cognitive capacity (Del Missier et al., 2013).

Behavioral and neurological studies have explored the relationship between working memory and reasoning (Buehner et al., 2005; Del Missier et al., 2013; Stevenson, Heiser, & Resing, 2013). For example, Mutter and Pliske (1996) compared the ability of detecting the contingency between the young adults and older adults, and found older adults were generally less accurate than young adults in judging event covariation. Similar findings were also evident in Mutter and Williams (2004), and Mutter, Strain, and Plumlee (2007). The authors suggested that the poorer performance of the older adults can be due to age-related decline in working memory capacity. Furthermore, the neuroimaging study conducted by Satpute et al. (2005) revealed that the left dorsolateral prefrontal cortex, of which the relationship with visual and verbal working memory has been evident in previous literature, had higher level of activation for the causal reasoning task than for the association judgment task. More recently, Braver, Burgess, Conway, and Gray (2011) found a strong correlation between reasoning test results and the level of activation in the postulated control mechanisms in WM tasks.

On the other hand, causal reasoning also depends on the quality of evidence received by reasoner. Uncertainty can be pervasive in real life situations, where the information is incomplete and the observations can be ambiguous. A recent neuroimaging study revealed that the activation areas when people were reacting to ambiguity and sample space ignorance, were overlapped with the areas associated with cognitive functions that are responsible for deliberate processing (Pushkarskaya, Liu, Smithson, & Joseph, 2010). This suggests reasoning and decision making under ambiguity may take up extra cognitive resources.

To date, little investigation has been undertaken in understanding how people reason causal relationships with ambiguous evidence, and little is known about the role of working memory in this process. The purpose of the current study is to investigate the relationship between working memory, causal reasoning and effects of ambiguous evidence.

WMC and Causal Reasoning

The framework of working memory proposed by Oberauer, Süß, Wilhelm, and Wittman (2003) states that the two major facets of working memory included a content facet and a function facet. A content facet is associated with the ability to store and maintain the information in the memory span over the time of the problem solving, while a functional facet is responsible for actively processing and manipulating the information. The functional facet covers a range of executive functions such as selection, inhibition, and interference resolution.

How does working memory function in causal reasoning? First, the span facets can be essential for data acquisition and retaining, in support of later hypothesis generation and inferences in causal reasoning (Lange, Thomas, Buttaccio, Illingworth, & Davelaar, 2013). WM capacity has been known to be associated with the depth of evidence encoding (Jaeggi, Buschkuehl, Perrig, & Meier, 2010), the recency effect (Sasaki, 2009), and the ability of retaining evidence that is rapidly presented (Lange et al., 2013). A higher span of WM should allow for more comprehensive evaluation of causal evidence.

Next, both span and the executive function facets can be associated with bias in evidence acquisition in causal reasoning. Mutter and Pliske (1996) found older adults were more likely to pay more attention to the instances where the cause is present, than the instances where the cause is absent. In addition, the older adults consistently provided positive causal judgments for the trials where the actual contingency was negative. This suggested that, when WM is limited, the data acquisition would become more selective, and might be more biased towards the instances where both cause and effect occur.

Furthermore, WM capacity can determine the maximum number of hypothetical causal structures that can be generated (Sprenger et al., 2011). De Neys, Schaeken, and D'Ydewalle (2005) found subjects with higher WM were less likely to accept logically invalid arguments such as denial of the antecedent, and affirmation of the consequent, than subjects with lower WM. De Neys et al. (2005) suggested that the tendency to accept these two invalid statements depend on the ability of the subjects to retrieve the counterexamples (including alternative and disabliers) in reasoning (De Neys et al., 2005). The number of hypothetical causal structures is determined by the number of alternative causes that are retrieved by the reasoner. In brief, previous evidence implies that WMC can influence the evidence acquisition and hypotheses generation in causal reasoning.

In the current study, we present subjects with causal evidence where we alter the covariance between the cause and the effect. If WMC is negatively associated with evidence selection bias, especially the tendency to rely on the positive evidence, we expect to observe subjects with lower WMC should have more positive causal ratings.

WMC and Ambiguity

Smithson, Bartos, and Takemura (2000) demonstrated that people are averse to ambiguity in decision making, and people are not ignoring the ambiguous information. Similarly, people may not merely omit the ambiguous observations in causal reasoning. This implies that the effects of ambiguity and uncertain outcomes on people's causal reasoning may not be the same as the effects of change in an observed sample quantity.

Next, Pushkarskaya et al. (2010) found that ambiguity and unknown outcomes are associated with multiple brain areas for deliberate processing. For example, there was a stronger activation in the inferior parietal lobe (IPL) when subjects were making decisions with ambiguous information and unknown outcomes, than they did in the risk situation. IPL was previously evident as a major area associated with sample space partitioning, evaluation of probabilities associated with each possible outcome, or integration of numerical information (Pushkarskaya et al., 2010). We have argued that WMC can be an indicator of a range of cognitive processing abilities, such as the depth of evidence encoding. This implies that subjects with higher WMC have higher capacity for the deliberately processing in dealing with ambiguous observations. In other words, we would expect to see an interaction between ambiguity and WMC, as subjects with higher WMC may be more likely be influenced by ambiguity due to the greater capacity to process the ambiguous information.

Methods

Participants and Design

A total of 72 subjects participated in the experiment to fulfill a course requirement. The experiment involved a causal reasoning task and two working memory tests. Two subjects did not successfully complete the two working memory tests, and were excluded from the data analysis. The remaining 70 subjects (44 females) had an average age of 21.33 years (*SD* = 4.21). The causal reasoning task contained 6 experimental blocks, which comprise the 2 (Ambiguous condition vs. Unambiguous condition) x 3 (contingency levels) within subject conditions. All tasks were computer based and were programmed in Inquisit 4.0.

Materials

Causal Reasoning Task Subjects were asked to pretend to be the employees of a neurovirology research institution, and their task was to evaluate the effects of a range of chemicals on certain types of viruses which cause neurological diseases. They would observe the paired stimuli that indicated the status of the virus and the presence or absence of the chemical. The status of a sample virus was either activated or inactivated. They would judge whether the chemical is an activator or inhibitor of the testing virus after observing a number of paired stimuli.

Each experimental block contained the 32 pairs of stimuli, which described the effects of one type of chemical on one type of virus. Each stimulus consisted of a picture of a virus on the right side of the screen, and information that indicated whether the virus had been or had not been exposed to the testing chemical. The color of the virus was either red (activated) or blue (inactivated). If the virus had been exposed to the chemical, a chemical structure picture was displayed on the left side of the screen. Otherwise, a capitalized text 'NO TESTING CHEMICAL' was displayed.

After observing the first 16 paired results, participants were asked to make their first judgment of the causal relationship between the virus and the chemical. They were asked "How likely do you think it is that the chemical activates or inactivates this type of virus?". They rated on a scale from -100 to 100. They were told that a negative rating means the chemical inactivates this type of virus; a positive rating means the chemical activates this type of virus; and a zero rating is appropriate when they think the chemical had no effects on the activation of this type of virus. They also rated their confidence on a 1 to 10 scale. They then proceeded to observe the remaining 16 paired stimuli. After observing all 32 sample viruses, they made their final judgment. Rated beliefs of the causal relationships, as well as confidence, were similar to the ones after 16 observations.

Both the unambiguous condition and the ambiguous condition had three experimental blocks. In the ambiguous condition, some outcome viruses were represented by a grayed image with a question mark in the center. The combination of the stimuli was shown in Table 1. The three blocks in each ambiguous/unambiguous condition section were randomly presented to subjects. The order of the ambiguous and unambiguous conditions were counter-balanced. The instruction before the ambiguous condition was that, "for some rea-

Table 1: Experimental Stimuli

Chemical	Present		Absent	
Virus	Inactive	Active	Inactive	Active
Contingency	Unambiguous Condition			
Positive	4	12	12	4
Zero	12	4	12	4
Negative	12	4	4	12
	Ambiguous Condition			
Positive	2 + 2A	10 + 2A	12	4
Zero	10 + 2A	2 + 2A	12	4
Negative	10 + 2A	2 + 2A	4	12

son, the status of some viruses were not clear, they were represented by the grayed picture with a question mark on it".

Single *N*-back Task The *n*-back test has been evident been associated with a wide range of reasoning abilities (Jaeggi, Studer-Luethi, et al., 2010). Attention, selection, inhibition and recognition are involved in the task (Chuderski & Necka, 2012). Subjects are required to hold a representation of an item in mind, while keeping the track on the current stimuli that need be compared with the stored representation, and actively updating the representation after each match takes place.

We applied the single *n*-back task developed by Jaeggi, Studer-Luethi, et al. (2010). In each experimental trial, subjects were presented a sequence of 20 letters. Subjects were required to press "A" on the keyboard when the current letter matched the one from n-steps earlier in the sequence. We tested 1-back, 2-back and 3-back. Subjects were instructed doing 3 practice trials (1 for each *n*-back) with feedback on their total proportion of correctness. After the practice trials, they completed 9 experimental trials: 3 trials for each *n*-back condition. The scoring of the *n*-back task depends on the total number of correct responses and the number of false alarm responses (subjects pressed the button when they should not do). The final score was calculated by the total number of hits subtracted by the total number of false alarms, averaged over the 9 blocks.

Operation Span Task The operational span (OSPAN) task requires subjects to remember a sequence of letters, while doing a number of mathematical problems. Focused attention, memory span and active arithmetic information processing were assessed in the task (Conway et al., 2005). A recent neuroimaging study has demonstrated that performing the Ospan task requires great activation in the IPL (Faraco et al., 2011). The Ospan task been demonstrated to have reasonable test-retest reliability, and good convergence validity with other WM measures and higher order cognition tasks (Conway et al., 2005).

We used the Ospan task developed by Unsworth, Heitz, Schrock, and Engle (2005). At the beginning of the task, subjects were instructed to go through 3 practice blocks: a block with the recalling task only, a block with math question solving only and a final block with doing both tasks. The experimental trials combined the two tasks. The length of letter sequence ranged from 3 to 7. There were three trials for each of these sequence lengths, and 15 trials with 75 letters in total. The final score used in later analysis was the total number of correctly recalled elements from trials in which all letters were recalled in correct serial order, and with more than 85% accuracy in mathematics .

Procedure

The experiment was conducted in a small group session format.The group size ranged from 5 to 12 subjects. Each subject worked on a computer individually. At the beginning of the experiment, the experimenter instructed the group of subjects. For the causal reasoning tasks, each subject was provided a work-sheet to write down their answers and notes regarding the chemical and viruses in each block. They were only allowed to write on the worksheets after they observed all stimuli on one block. Subjects firstly went through a self-paced practice block, in which they pressed the space-bar to proceed with the paired stimuli. Then they completed the six experimental blocks, where each paired stimuli stayed on screen for two seconds, and automatically proceeded to the next pair. After completing the causal reasoning task, participants were required to take a 5-10 minutes break. Then they proceeded to the two working memory tasks. There was an approximate 5-10 minutes break between the two working memory task.

Results

The mean *n*-back score was 3.91 (SD = 0.66, median = 1.00), and the mean Ospan score was 45.59 (SD =15.55, median = 44.5). There was a moderate positive relationship between the *n*-back scores and Ospan scores, $r = 0.39$, $p = .009$. A negative power transformation was applied to the *n*-back scores due to its negative skewness. Both the Ospan and transformed *n*-back scored were standardized in later analysis. Figure 1 and 2 show the descriptive results of causal ratings.

Beta regression analysis with Bayesian parameter estimation was applied to examine the effects of experimental condition, and working memory on both the mean and the variability of subjects' causal estimates. We used ΔDIC – the contribution of a factor to the model DIC (Deviance Information Criterion) – to assess the effects of different factors. A positive ΔDIC value indicates that a model with that factor has better DIC fit values than a model without that factor. A ΔDIC value over 3 indicates a substantial contribution of a factor. We reported the estimated parameter and 95% credibility interval (95%CI) for the individual factor in either the location submodel or the precision submodel of the beta regressions. A positive coefficient value in the location submodel indicates a positive relationship between the factor and the mean of the dependent variable (e.g., causal ratings). A positive coefficient value in the dispersion submodel indicates a positive relationship between the factor and the precision (or homogeneity) of the dependent variable. Lower precision of the dependent variable suggests greater individual differences in providing the responses[1].

The Effects of Observation Quantity

First, we examine the effects of sample observation quantity on the causal ratings. The causal ratings after 16 observations, and the ratings after 32 observations in the unambiguous condition were included in the analysis. Contingency had significant main effects on the the mean and the dispersion of the causal rating (ΔDIC = 226.1, and 19.25 for the contribution of contingency to the location and dispersion submodel, respectively). Post-hoc repeated comparisons showed that the ratings of the positive condition were significantly higher than the zero contingency condition (b = 0.77, 95%CI = [0.59, 0.95]), and the ratings of the zero condition were significantly higher than the negative condition (b = 1.09, 95%CI = [0.89, 1.31]). It was also found that the mean ratings in the zero condition were significantly greater than 0 (b = 0.32, 95%CI = [0.20, 0.45]).

The ratings of the negative contingency condition had significantly greater variability than the positive (d = -0.60, 95%CI [-0.92, -0.30]) or the zero contingency conditions (d = -0.72, 95%CI = [-1.04, -0.41]). There was no significant difference between the positive contingency and the zero contingency conditions in the variability of the causal ratings (d = 0.09, 95%CI [-0.21, 0.41]).

The number of observations did not have a significant contribution to either the mean or the dispersion of the causal ratings (ΔDICs $<$ 3), suggesting that there was no significant difference between the causal ratings after 16 observations, and the causal ratings after 32 observations. *N*-back did not significantly predict the mean causal ratings as a main factor (ΔDIC = -2.2); however, it was a significant predictor of the variability of the causal ratings (ΔDIC = 12.6; d = 0.27, 95%CI = [0.12, 0.42]). Subjects with higher *n*-back scores provided more homogeneous causal ratings than those who had lower *n*-back scores. Finally, Ospan scores did not have significant relationship with either the mean or the dispersion of causal ratings (ΔDICs $<$ 3). No significant interactions were found between sample quantity and the two working memory measures.

Confidence There was a significant effect of contingency on the mean and dispersion of subjects' confidence ratings (ΔDIC = 29, and 42 for the location and dispersion submodel, respectively). Post hoc repeated comparisons revealed that subjects had significantly lower confidence ratings in the negative contingency condition than the positive condition (b = -0.17, 95%CI = [-0.34, -0.01]). The confidence ratings of both the zero and the negative contingency conditions were significantly more variable (i.e., less homogeneous) than the positive contingency condition (d = -0.92, 95%CI = [-1.30, -0.54] for comparing the zero contingency condition with the positive contingency condition; d = -0.68, 95%CI = [-0.33, -0.05] for comparing the negative contingency condition with the positive contingency condition).

The observation quantity and *n*-back did not have significant effects on either the mean or the dispersion of the confidence ratings of subjects (ΔDICs $<$ 3). There was a substantial effect of Ospan on the dispersion of the confidence ratings. There was greater variability in confidence ratings among subjects with higher Ospan scores (d = -1.07, 95%CI = [-0.33, -0.27]).

[1]More details about beta regression, as well as experimental stimuli are available in the supplementary material at http://goo.gl/LSnJNN

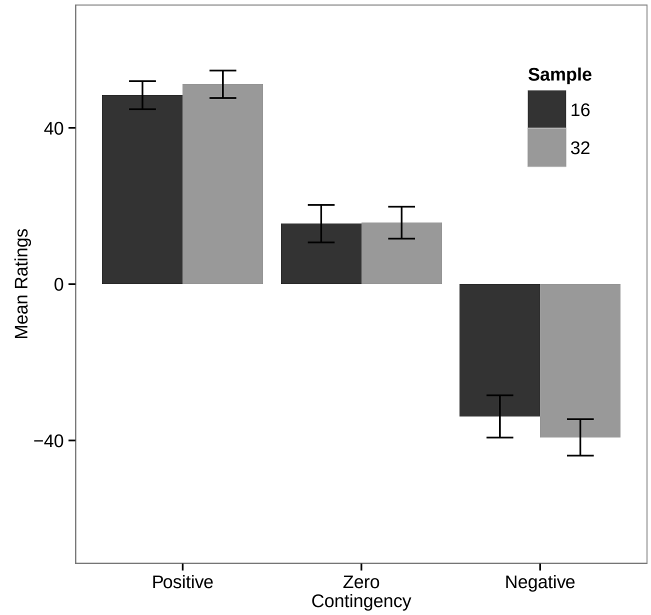

Figure 1: Mean Causal Ratings after 16 Observations vs after 32 Observations

The Effects of Ambiguity

The causal ratings after 32 observations for both the unambiguous and ambiguous conditions were analyzed by using beta regressions. The effects of contingency were similar to the results in the previous section: it had significant main effects on both the mean and the precision of the rating (ΔDIC = 209.79, and 19.54 for the location and precision submodel, respectively). The ratings of the positive contingency condition were significantly higher than the zero contingency condition (b = 0.77, 95%CI = [0.59, 0.95]), while the zero contingency condition had significant higher mean ratings than the negative contingency condition (b = 1.09, 95%CI = [0.89, 1.31]). There was no significant difference between the positive contingency and the zero contingency conditions in the dispersion of causal ratings (d = 0.09, 95%CI = [-0.21, 0.41]). The ratings of the negative contingency condition were significantly more heterogeneous than either the positive contingency condition (d = -0.60, 95%CI = [-0.92, -0.30]) or the zero contingency condition (d = -0.72, 95%CI = [-1.04, -0.41]).

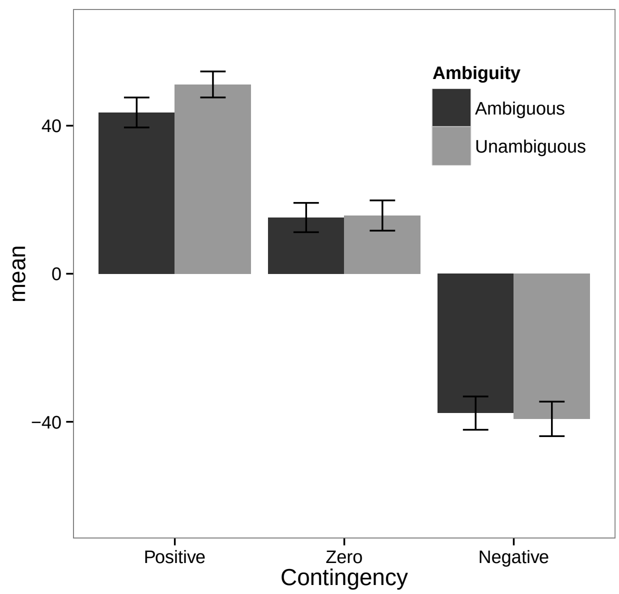

Figure 2: Mean Causal Ratings of the Unambiguous and Ambiguous conditions

Ambiguity and the two working memory measures did not have significant main effects on either the mean or the precision of the ratings (ΔDICs < 3). However, there was a significant interaction between Ospan and ambiguity on the variability of the causal ratings (ΔDIC = 7.1). The higher Ospan scores were associated with more homogeneous causal ratings in the unambiguous conditions (d = 0.31). In contrast, the higher Ospan scores were associated with more heterogeneous causal ratings in the ambiguous condition (d = -0.11).

Finally, a significant interaction was found between n-back and contingency conditions on the mean of the causal ratings (ΔDIC = 5.4). Post-hoc repeated comparisons revealed that n-back did not have significant main effects on the mean causal ratings for the zero and negative contingency conditions, however, n-back scores had a significant positive association with the positive contingency condition (b =0.16, 95%CI = [0.02, 0.29]).

Confidence Contingency had significant main effects on both the mean and the dispersion of the rating (ΔDIC = 24.4, and 32.8 for the factor contribution to the location and precision submodel, respectively).

Both ambiguity and n-back did not have significant effects on either the mean or the dispersion of the ratings (ΔDICs < 3). There was a significant interaction between n-back and ambiguity (ΔDIC = 4.9; d = 0.28, 95%CI = [0.11, 0.45]). N-back and the homogeneity of the confidence ratings had a significant positive relationship in the ambiguous condition, but not in the unambiguous condition. Subjects with higher n-back scores provided more homogeneous confidence ratings in the ambiguous conditions.

A significant interaction was found between ambiguity and contingency conditions in the precision submodel (ΔDIC = 11.6). Subjects provided more homogeneous confidence ratings in the zero contingency condition when there were ambiguous outcomes.

Finally, Ospan scores had a substantial effect on the precision of confidence ratings (ΔDIC = 9.3). There was greater individual difference in confidence ratings among subjects with higher Ospan scores (d = -0.23, 95%CI = [-0.38, -0.08]).

Discussion

There was no statistical evidence that WMC was associated with the tendency to have more positive estimates of the causal relationships. However, there was a significant relationship between n-back and the variability of the causal ratings of subjects. Subjects with higher n-back scores had more homogeneous ratings than those with lower working memory capacity.

N-back has been previously demonstrated to be related to inhibitory ability (Jaeggi, Buschkuehl, et al., 2010). Subjects with lower n-back scores may not be able to inhibit information from different sources (such as prior beliefs) as well as those with higher n-back scores, when keeping track on the current observations (Mutter et al., 2007). In addition, the n-back test involves assessing the active information updating process. Thus low n-back scores may indicate poor ability in updating one's beliefs. Subjects with poorer inhibition ability may provide casual judgments that rely more on their prior beliefs. The individual differences in the prior beliefs regarding the association between the cause and effect in the current chemical-virus scenario, consequently, can contribute to the higher variability in causal ratings among subjects with lower n-back scores.

There was no significant difference between causal ratings after 16 observations and their ratings after 32 observations. In addition, these results were not dependent on the performance of the two WMC measures. This suggests that the stabilization of causal beliefs may not depend on WMC in the current study. If subjects simply ignored the ambiguous information, which was only 12.5% of the observations, we should expect to observe similar ratings patterns of subjects between the ambiguous and unambiguous conditions.

Turning now to the experimental evidence on the effects of ambiguity, we found a significant interaction between Ospan and ambiguity. There was a greater variability in causal ratings among subjects with higher Ospan in the ambiguous condition than those who had lower Ospan. It has mentioned earlier that both Ospan and ambiguity was associated with the activation in the brain areas that are associated with evaluation of probabilities related to each possible outcome and arithmetic operations (Faraco et al., 2011; Pushkarskaya et al., 2010). These findings, and the current results imply that dealing with ambiguity in causal reasoning can be associated with arithmetic computation and probability evaluation.

Pushkarskaya et al. (2010) observed that the activation of IPL is moderated by individual differences in tolerance of ambiguity. Subjects who were more ambiguity averse had greater activation than those who were less ambiguity averse. This suggested that the subjects who were more ambiguity averse were more likely to employ deliberative processing in the ambiguous condition. Ospan assesses the ability to temporarily retain information, and actively process arithmetic information simultaneously, and therefore can be a reasonable indicator of the individual's capacity in deliberative processing. Therefore, individual differences in tolerance of ambiguity may contribute to the greater dispersion of the causal ratings among subjects with higher Ospan in the current study. On the other hand, subjects with lower Ospan had smaller variation in causal ratings due to limited available capacity.

In summary, the current study provides a preliminary investigation of the relationship between working memory, causal reasoning and ambiguity. The findings suggest that subjects were not merely ignoring the ambiguous information in causal reasoning, and the way in which one processes the ambiguous information can be associated with one's working memory capacity. A more comprehensive investigation may require the consideration of the individual differences in attitudes toward ambiguity, as well as a sample with a greater variability in WMC.

References

Braver, T. S., Burgess, G. C., Conway, A. R. A., & Gray, J. R. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General*, *140*, 674–692.

Buehner, M., Krumm, S., & Pick, M. (2005). Reasoning=working memory not= attention. *Intelligence*, *33*(3), 251–272.

Chuderski, A., & Necka, E. (2012). The contribution of working memory to fluid reasoning: Capacity, control, or both? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1689–1710.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786.

De Neys, W., Schaeken, W., & D'Ydewalle, G. (2005). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning*, *11*(4), 349–381.

Del Missier, F., Mäntylä, T., Hansson, P., Bruine de Bruin, W., Parker, A. M., & Nilsson, L.-G. (2013). The multifold relationship between memory and decision making: An individual-differences study. *Journal of experimental psychology. Learning, memory, and cognition*, *39*, 1344–64.

Faraco, C. C., Unsworth, N., Langley, J., Terry, D., Li, K., Zhang, D., … Miller, L. S. (2011). Complex span tasks and hippocampal recruitment during working memory. *NeuroImage*, *55*, 773–787.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory (Hove, England)*, *18*(4), 394–412.

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning implications for training and transfer. *Intelligence*, *38*(6), 625–635.

Lange, N. D., Thomas, R. P., Buttaccio, D. R., Illingworth, D. A., & Davelaar, E. J. (2013). Working memory dynamics bias the generation of beliefs: the influence of data presentation rate on hypothesis generation. *Psychonomic bulletin & review*, *20*(1), 171–6.

Mutter, S. A., & Pliske, R. M. (1996). Judging Event Covariation: Effects of Age and Memory Demand. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *51B*(2), P70–P80.

Mutter, S. A., Strain, L. M., & Plumlee, L. F. (2007). The role of age and prior beliefs in contingency judgment. *Memory & Cognition*, *35*(5), 875–884.

Mutter, S. A., & Williams, T. W. (2004). Aging and the detection of contingency in causal learning. *Psychology and Aging*, *19*(1), 13–26.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory. *Intelligence*, *31*(2), 167–193.

Pushkarskaya, H., Liu, X., Smithson, M., & Joseph, J. E. (2010). Beyond risk and ambiguity: deciding under ignorance. *Cognitive, affective & behavioral neuroscience*, *10*, 382–391.

Sasaki, T. (2009). Individual differences in working memory capacity in recency effects: from the recall process. *Psychological reports*, *104*(2), 545–8.

Satpute, A. B., Fenker, D. B., Waldmann, M. R., Tabibnia, G., Holyoak, K. J., & Lieberman, M. D. (2005). An fMRI study of causal judgments. *The European journal of neuroscience*, *22*(5), 1233–8.

Smithson, M. J., Bartos, T., & Takemura, K. (2000). Human judgment under sample space ignorance. *Risk Decision and Policy*, *5*(02), 135–150.

Sprenger, A. M., Dougherty, M. R., Atkins, S. M., Franco-Watkins, A. M., Thomas, R. P., Lange, N., & Abbs, B. (2011). Implications of Cognitive Load for Hypothesis Generation and Probability Judgment. *Frontiers in Psychology*, *2*, 15.

Stevenson, C. E., Heiser, W. J., & Resing, W. C. (2013). Working memory as a moderator of training and transfer of analogical reasoning in children. *Contemporary Educational Psychology*, *38*(3), 159–169.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior research methods*, *37*(3), 498–505.