The Relations between Time, Confidence and Accuracy: Using a Judgment Questionnaire

Soon-Li Lee
School of Medicine and Health Sciences
Monash University Sunway Campus
Malaysia.

Cai-Lian Tam
School of Medicine and Health Sciences
Monash University Sunway Campus,
Malaysia

Teck-Heang Lee
Taylor's Business School
Taylor's University
Malaysia.

Wai-Mun Har
Faculty of Accountancy & Management
Universiti Tunku Abdul Rahman
Malaysia

Abstract

This study aims to explore the interaction between confidence of judgment, accuracy and time. 86 university students participated in the study. 43 participants were assigned to the fast condition. They were required to answer a question within six seconds and another six seconds were provided to rate their level of confidence for their response. The remaining participants were assigned to the slow condition and repeated the same procedures with nine seconds. Present study revealed that the participants' confidence rate was not equal to the actual probability of being correct. Furthermore, there was no relationship between accuracy and time for deliberation. The results failed to support the hypothesis that when more time is provided, both participants' responses accuracy and confidence rate will increase. Finally, the prediction that accuracy should increase with higher ratings of confidence was supported. It was concluded that time has very little effect on calibration of judgment processes.

Keywords: Time, accuracy, judgment

1. Introduction

It was widely believed that confidence was an adequate predictor of accuracy. However, the way in which confidence is assigned was largely an imperfect process and reliant on one's perception and feelings over currently concrete or definable parameters (Deffenbacher, 1980). The judicial system consistently relied on confidence as a primary indicator or perceived accuracy in eyewitness testimonies, so understandably, knowledge of how confidence attribution worked was imperative to understanding how accuracy can be measured and impacts on resulting judgments (Wells, 1984).

The variables involved in the present study were confidence, overconfidence, underconfidence, accuracy, and calibration. Previous researchers defined confidence as the probability of a judgment for being correct (Paese & Sniezek, 1991). Overconfident is when a judgment is believed to be more accurate than it should be; whereas underconfident is when a judgment is believed to be less accurate than it is. Normally people tend to be overconfident in their judgment (Lichtenstein & Fischhoff, 1977). Lord (1985) explained that accuracy occurred when a person is able to effectively process information and distinguishes it from other information which seems to be plausible from their knowledge. A judgment is calibrated if the probability of judgments matches with the corresponding relative frequency of occurrence (Brenner, Koehler, Liberman & Tversky, 1996). In addition, if decision makers know how accurate they are, they are well calibrated (Lichtenstein & Fischhoff, 1977).

An organization known as the Innocence Project was established in order to assist individuals with wrongly convicted crimes. In 2008, The Innocence Project assisted 220 wrongly accused criminals, which were more than two-thirds compared with the amount of assistance provided in 2000. Eyewitness with high level of confidence in reporting their memories tend to be more persuasive. However, this event suggested that eyewitness' statements might not be reliable in court setting (Holmes & Weaver III, 2010). This addressed a need to conduct further research in this area to support or to contradict the usefulness of eyewitnesses in court setting. It is crucial as the reliability of eyewitness might determine the outcome of the court's decision.

2. Literature Review

2.1 Confidence vs accuracy

A research on accuracy and confidence was conducted by Ordinot, Wolters and van Koppen (2009) using 14 witnesses for an armed robbery. The study revealed that most of the recalled information with high level of confidence was accurate but the information with the highest level of confidence was found incorrect. This research concluded that confidence may be used as a cautious indicator for accuracy during police investigation but it cannot be used as evidence. In addition, there was no standardization on confidence level, unable to determine the level of confidence which accompanied participants' responses. Ames, Kammrath, Suppes, and Bolger (2010) conducted a study using 25 university students. They found that different groups of participants were required to make impression judgments accompanied with their confidence level towards their judgment. Overall, they found that accuracy does not regularly predict confidence. This study acknowledged that there is a possibility for accuracy and calibration differ with different materials and participants in study.

Ordinot, Wolters and van Koppen (2009) and Ames, Kammrath, Suppes, and Bolger (2010) argued that confidence level is not a reliable predictor of accuracy even though there are other researchers provided contradicted results to their arguments. Among other, include, Kebbell (2009) who had conducted a research on witness' confidence and accuracy during an interview with 100 participants. This experiment concluded that participants were more confident about correct answers, as they are able to judge answers' accuracy in a precise manner. This study found that with the presence of interviewer, participants' confidence will increase. However, interviewers tend to rates witness as less confident then witness themselves.

A study conducted by Mengelkamp and Bannert (2010) which required their participants to learn about the principles and the application of operant conditioning as well as The knowledge towards the learned materials of the participants was tested with several tests. During the tests, participants were required to rate their level of confidence for their responses. The analysis revealed that participants, before learning, provided low judgment for their level of accuracy. However, a higher judgment of accuracy was given after learning the materials. Eventually, participants reported accurate judgments for all tests, suggesting that participants were able to judge their knowledge better than chance. They found that stability in absolute accuracy was significant. Moreover, they found that relative accuracy is domain specific, allowing it to generalize within a domain only. Their third hypothesis was supported due to certain constraints imposed by model used in the experiment. In a similar vein, Gonzalez-Valejjo and Bonham (2007) conducted a research to study the relationship between feedback and accuracy. This study supported that increased accuracy can lead to improvement of calibration as overconfidence diminished as accuracy improved. They also found that accuracy and confidence increase with greater repetition rate. Confidence level tends to increase tremendously for items answered correctly, while different pattern can be observed for items answered incorrectly, where confidence level slightly increased.

2.2 Confidence vs time

Vickers and Packer (1982) examined the relationship between response time and confidence when accuracy was varied across trials. This study involved twenty participants. The study found, under the condition which demands for accuracy, participants generally produced more correct responses, required longer response time, and made higher confidence rating compared with participants' in time demanding condition. This study concluded the existence of an inverse relationship between confidence rating and response time. Besides, high confidence rating was accompanied with shorter response time or vice versa was noted. A study to investigate the relationship between decision times, subjective probability, and task difficulty using probability assessment task was conducted by Wright and Ayton (1988). Their participants were required to choose the perceived accurate answers and rate their confidence level for each answer's accuracy with the participants' decision time being recorded. It was found that decision time increased as task difficulty increases. They also found that decision time increased as overall proportion of correct responses increase.

Another research conducted to examine the relationship between time and accuracy was carried out by Dunning and Perretta (2002). Their research design required different group of participants to engage in different type of identification tasks with a record of their decisions and confidence level. This study found that response made within 10-12 second will be more accurate. However, if the participants' responses were automatic, it should consume lesser time.Brewer, Caon, Todd and Weber (2006) conducted experiments which examined the relationship of time and accuracy. Brewer et al (2006) argued that optimum time boundary increased with the lineup size, which is task load, and mean response latency or the decision time. Brewer found the optimum time boundary varied from 5 to 36 seconds and their findings are not in line with the earlier study of Dunning and Perretta's (2002) which found that the accurate responses are made within 10-12 seconds. Pleskac and Busemeyer (2010) research predicted that an increase in *t* values will lead to increase in mean differences and the pooled variances of confidence ratings in correct and incorrect responses. This study revealed that time has only moderate effect on accuracy and confidence as people tend to minimize decision time to maximize their choice and confidence accuracy.

Ordinot, Wolters and van Koppen's (2009) study used witnesses for an event which was 3 months away from the research. They found that there were some details which was not accurate, suggested the effect of time. Kebbell (2009) suggested future researchers to use different scale of difficulties as the conducted experiment might not be applicable as in real life setting questions difficulty might not be the same with the experimental questions. Ames, Kammrath, Suppes, and Bolger (2010) acknowledged that urgency might influence their study result. Although Gonzalez-Valejjo and Bonham's (2007) study showed that familiar material can influence calibration, their study was not enough to explain the different in confidence and accuracy alignments.

Time is an important variable that can be used to explore the effect of deliberation on accuracy and confidence. Vickers and Packer (1982) support this by showing how when observers take longer to respond to a question and less errors were made. Wright and Ayton (1988) also concluded that the more time allowed for deliberation positively correlated to increased accuracy of response. Vickers and Packer (1982) however found a contradictory result that suggested time was inversely related to confidence, so that the less time taken to make a judgment is equated with a larger likelihood of a higher confidence reporting. Vickers and Packer (1982) also notes that this idea is counter intuitive and further research needs to be conducted in the area. Dunning and Perretta (2002) could not find the time frame which participants consistently commit error, suggesting that this study should use a different time frame from them. Brewer, Caon, Todd and Weber (2006) found the strong relationship between accuracy and response latency, but they acknowledged that more studies should be conducted as it is still too early to conclude anything. Pleskac and Busemeyer's (2010) study failed to answer how to different types of time pressure influence accuracy and their application in everyday life.

The aim of this study is to test whether calibration is influenced by the time available for deliberation. The independent variables involved in this study were speed of presentation of questions, and the level of confidence for each answer. The dependent variable was the percent of correct answer. The reason for doing this research is to extend the finding from previous study, which did not explain the relationship of time, accuracy, confidence and calibration. There were four hypothesis involved in this study.

H1: The participants' confidence rate is equal to the actual probability of being correct.

H2: A highly accurate participants' responses should accompany with high rating of confidence.

H3: With more time, there will be an increase in accuracy for participants' responses.

H4: With more time provided, which is in the slow condition, there will be an increase in participants' responses accuracy with higher confidence rate.

3. Method

3.1 Participants

86 students from an Australian university participated in the study. It comprises of 24 male and 62 female with the mean age of 22.09. They were selected on the basis of convenience sampling. The participants were assigned to experimental conditions on the basis of two psychology lab classes. The lab classes were randomly assigned to fast or slow conditions. Participants were not naïve to the study as they were required to familiarize themselves with the rating of confidence level and the true or false quiz at the required speed. Furthermore, the participants signed and handed their inform consent in their first tutorial class. After the practice, participants were required to conduct the actual questionnaire. Half of the participants were assigned to the slow condition, while another half was assigned to the fast condition.

3.2 Materials

The materials involved were questionnaire with 100 general knowledge questions of varying difficulty. This questionnaire was specially developed for this research. 86 OPSCAN sheets were required for scoring purpose. These OPSCAN sheets were commonly used for multiple choice examinations. It has the capacity of 200 responses, which make it necessary for this study as participants were required to provide their level of confidence for each response, which required 200 spaces in their response sheets. A stopwatch was used to measure time for slow condition and fast conditions.

3.3 Procedure

Participants were assigned to different conditions. 43 participants were assigned to the fast condition, where participants were required to answer a question within six seconds and another six seconds were provided to rate their level of confidence for their response. The remaining 43 participants were assigned to the slow condition, where participants were provided with nine seconds to provide their response for a question and another nine seconds to rate their level of confidence for their response. Participants were required to express their confidence in their responses' accuracy as a probability ranging from 0.55 as pure guessing, 0.65 as almost guessing, 0.75 as unsure, 0.85 as almost certain, and 0.95 as absolutely certain. Participants were required to familiarize themselves with the measurement of confidence by completing a set of practice with 10 questions. After participants were familiarized with the measurements of confidence, they were required to conduct the actual measurement by alternately answering and rating confidence on the OPSCAN sheets. For this study, there were two independent variables, which were time or the speed of presentation of questions and the level of confidence, and one independent variable, which was the participants' accuracy for each response. The independent variables were manipulated by varying the speed of presentation of questions, which were six second for fast condition and a slow condition with nine second, and by varying the level of confidence with 0.55 as pure guessing, 0.65 as almost guessing, 0.75 as unsure, 0.85 as almost certain, and 0.95 as absolutely certain. The dependent variable, which was participants' accuracy, was measured by calculating the percentage of correct responses from the participants.

4. Results

A 2 x 5 mixed model ANOVA was conducted to analyze the impact of two experimental conditions, which were the fast and slow condition, on the level of accuracy for participants own responses. This model of ANOVA was chosen as the data were obtained through repeated measures on participants from two groups. A series of dependent sample *t* tests were used to calculate whether accuracy increase with the increase in confidence. An Independent Samples *t* Tests was conducted to detect difference in accuracy between fast and slow conditions at each level of confidence. A series of One Sample *t* Tests were conducted to explain the calibration. The data were arranged into five different categories, corresponding to the five confidence ratings. The raw data served as difference data among five confidence ratings. The summary of means for fast and slow group and standard deviations for levels of confidence for ANOVA were summarized in Table 1.

Table 1. Means for Fast and Slow Condition and Standard Deviations for Level of Confidence for ANOVA

	Fast		Slow	
Level of confidence	Means	*SD*	Means	*SD*
Guess	47.40	11.65	50.10	10.20
Almost guess	53.34	16.80	51.99	14.59
Unsure	54.60	14.50	55.00	15.30
Almost certain	62.20	15.10	62.73	14.10
Certain	70.20	12.76	68.23	13.51

From ANOVA, there was a significant main effect of confidence, with the F (4, 336) = 30.03, $p < .05$, with a large effect of $\eta^2 = .26$. Statistically, there was a significant difference between accuracy and confidence. A series of dependent sample *t* tests compared the guess (M = 48.74, SD = 10.96) with almost guess (M = 52.67, SD = 15.64) with the result of t (85) = -2.07, p <.05, with a moderate effect of r^2 = 0.05, guess with unsure level (M = 54.80, SD = 14.80) with the result of t (85) = -2.76, p <.05, with moderate effect of r^2 = 0.08, guess with almost certain level (M = 62.45, SD = 14.09) with the result of t (85) = -6.88, p <.05, with a large effect of r^2= 0.36, guessing level with certain level (M = 69.20, SD = 13.10) with the result of t (85) = -11.15, p <.05, with a large effect of r^2= 0.59.

This *t* tests revealed that there was an increase in accuracy with the increase of confidence. From ANOVA, the main effect for both experimental groups was not significant, F (1, 84) = .003, $p > .05$, $\eta^2 = .00$. Statistically, participants reported same performance in slow and fast condition. From the same ANOVA, there was no significant interaction between experimental conditions and confidence, F (4, 336) = .0.37, $p > .05$, with a small effect of $\eta^2 = .004$. Statistically, there was no extra mean difference which was not explained by the main effects of confidence and experimental conditions. Table 2 provide a summary of means and standard deviation for all level of confidence in slow and fast conditions, *t* values, *p* values, and effect size for Independent Samples *t* Tests.

Table 2. The Means and Standard Deviations for All Level of Confidence In Slow and Fast Conditions, t values, p values, and effect size for Independent Samples t Tests

Level of confidence	Fast		Slow				
	Mean	*SD*	Mean	*SD*	*t*	*p*	r^2
Guess	47.40	11.65	50.08	10.18	-1.13	-.26	0.02
Almost guess	53.34	16.78	51.98	14.59	-0.39	-.69	0.00
Unsure	54.60	14.50	54.99	15.27	-0.12	-.90	0.00
Almost certain	62.17	15.10	62.73	13.20	-0.18	-.85	0.00
Certain	70.17	12.76	68.22	13.51	-0.69	-.49	0.01

From Table 2, the non significant Independent Samples *t* Tests with the degree of freedom of 84 supported that there was no difference in accuracy between fast and slow conditions at each level of confidence, with small effect. A series of one sample *t* tests was conducted. For the guess level (M = 48.74, SD = 10.96), the reported *t* is t (85) = -5.29, $p < .05$, with $r^2 = 0.25$ which indicated large effect, with the used test value of 55. For the almost guess level (M = 52.67, SD = 15.64) reported value of t (85) = -7.31, $p < .05$, with the $r^2 = 0.39$ which indicated large effect, with the used test value of 65. The unsure level (M = 54.80, SD = 14.80) reported t (85) = -12.66, $p < .05$, with the $r^2 = 0.65$ which indicated large effect, with the used test value of 75. The almost certain level (M = 62.45, SD = 14.09) reported t (85) = -14.85, $p < .05$, with the $r^2 = 0.72$ indicated large effect with the used test value of 85. The certain level (M = 69.20, SD = 13.10) with the result of t (85) = -18.27, $p < .05$, with $r^2 = 0.79$ suggested a large effect, with the used test value of 95. This supported that participants were not perfectly calibrated, suggesting overconfident as their responses were not as accurate as they thought, with the means lower than test values.

5. Discussion

There were four hypotheses tested in this study. The study rejected the first hypothesis which suggested that participants' confidence rate is equal to the actual probability of being correct. However, the second hypothesis stated highly accurate participants' responses should accompany with high rating of confidence, was supported. The third hypothesis stated that with more time, there will be an increase in accuracy for participants' responses, was rejected. Lastly, the fourth hypothesis which stated that with more time provided, which is in the slow condition, there will be an increase in participants' responses accuracy with higher confidence rate, was also rejected. The first hypothesis, stated that participants' confidence rate is equal to the actual probability of being correct was not supported by this study, contradicted with Gonzalez-Valejjo and Bonham's (2007) finding, which is increased accuracy can lead to improvement of calibration as overconfidence diminished as accuracy improved. The current study found that participants tend to be overconfident in this study, which is similar with Lichtenstein and Fischhoff's (1977) finding, although accuracy improved with increasing confidence.

The reasons for such a result in this study could be due to the availability of information as the questions were testing participants' general knowledge, the varied difficulties of questions, and the absence of feedback for responses provided by participants (Speir-Bridge et al., 2010). In addition, participants who were consider as adults have the tendency to falsely combine the features of different sources, which contributed to high-confidence errors in their responses. Thus, most of the committed errors were caused by retrieval difficulty as the participants unable to specify the sources of the familiarity signals at retrieval (Shing, Werkle-Bergner, Li & Linderberger, 2009). The effect of overconfident can be produced by the presence of group work. Although participants worked individually, there is no prohibition posed by researchers and hence it is likely for them to work in group of two. By doing so, it improves the accuracy and increase the participants' confidence rate but it may bring about undesirable effect of overconfident (Puncochar & Fox, 2004).

The second hypothesis which stated that an increase in responses' accuracy should accompany with high rating of confidence, was supported. This finding is consistent with previous study in the literature for example: (i) Kebbell's (2009) finding where participants were able to judge answers' accuracy in a precise manner, and (ii) Mengelkamp and Bannert (2010), with the finding of participants were able to judge their accuracy better than chance. However, the results of the second hypothesis contradicted with Ames, Kammrath, Suppes, and Bolger's (2010) who found that accuracy does not regularly predict confidence. The contradictory results may can be explanation by Brewer and Sampaio (2006) who agued that the availability of the appraised items play an important role as participants tend to provide a higher rating for responses which are more familiar compare with the rest. Brewer, Sampaio and Barlow (2004) also asserted that the source of confidence is originated from memory accuracy. When their participants believed that they made a complete recall, their responses were likely to be accurate. Similarly, Sauerland and Sporer (2009) suggested a briefing before formal assessment allows participants to make better estimation of confidence. This explains the result from the study as participants were not naïve to the actual purpose of current study, and they were exposed to a trial in order familiarize the participants to the confidence scale. Another possible explanation why the second hypothesis was supported may due to group work. In the present study, it is possible for participants to discuss among themselves since there is no prohibition imposed by the researchers. Puncochar and Fox's (2004) argued that participants who worked in a group displayed higher confidence and accuracy in their decisions compare with participant who worked alone.

The third hypothesis stated that there should be an increase in responses' accuracy with more time, and the fourth hypothesis, stated that with more time there should be an increase in responses' accuracy with increasing confidence for the slow condition. Both hypotheses were not supported, partially validated with Pleskac and Busemeyer's (2010) finding, revealed that time has only moderate effect on accuracy. These findings validated Brewer, Caon, Todd and Weber's (2006) findings, suggesting that time is not practical to identify accuracy. However, the results of the present study contradicted with Dunning and Perretta's (2002) finding as they suggested that responses which being provided within 10 till 12 seconds are accurate compared with responses provided within a longer time range. Furthermore, Vickers and Packer's (1982) result was not supported by this study as they found that their participants committed fewer error when more time was provided.

Additionally, Wright and Ayton's (1988) study with the result which suggested more time allowed for deliberation positively correlated to increased accuracy of response was not supported as well. The short decision time can be explains by the presence of high post-decision confidence. Participants might perceive their responses were made by automatic processes, thus rated highly on their own responses (Sauerland & Sporer, 2007). Another possible explanation is the questions were straightforward, possibly constructed from simple phrases. When participants were provided with simple questions, a swift response is likely to be accurate. In contrary, for a complex question, a swift response might not be reliable (Kebbell, Evans, & Johnson, 2010). The fourth hypothesis which failed to be supported can be explained by the poor monitoring after task overload in participants (Boekaerts & Rozendaal, 2010). The overload could be caused by the long duration for participants to consider the best answers. Furthermore, the participants were dealing with 200 items which were cognitively demanding, added on their cognitive burden. Moreover, the long delay between the test and confidence rating led to overconfidence, as according to Van Overschelde and Nelson (2006). The decline of accuracy despite longer time provided determined by the nature of the task. Task which requires participants to recall will suffer from a decline in performance when there is a delay, as found by Koriat, Bjork, Sheffer and Bar (2004). The questionnaire developed for this study can be considered as a recall task because it is testing participants' existing knowledge on Australia culture. Thus, it led to underperformance in the slow condition which affected calibration, caused discrepancy between participants' accuracy and confidence rating.

This research failed to support the notion that with more time people can make more accurate judgment. This in turn suggested the application of participants' schemas or heuristics in completing the questionnaire. Paul and Nazareth (2010) supported this finding with the explanation that people tend to use schema when the input of information is too complex and when they were given limited time. These two factors contributed to overload, and the application of appropriate decision aid, which means schema can solve the problem of overload. It is possible for participants to be overloaded as they were dealing with 200 items. Participants in the slow condition might experienced more burden from the longer time to decide, which supported by Van Overschelde and Nelson (2006). However, the valid second hypothesis suggested that confidence can be a useful cue towards accuracy, as supported by Ordinot, Wolters and van Koppen (2009), Kebbell (2009) and Mengelkamp and Bannert (2010).

This study failed to support the importance of time, contradicted with Pleskac and Busemeyer (2010), Brewer, Caon, Todd and Weber (2006), Dunning and Perretta (2002), Vickers and Packer (1982) and Wright and Ayton's (1988) studies. Possible confounding which serves as limitations could be the relevance of questions in the questionnaire and the varying difficulty levels. Participants might not be able to relate well to the questions, leading to underperformance. According to Kebbell, Evans, and Johnson (2010), overconfident is likely to occur if the participants were unable to relate well with the questions. Thus, researchers may need to include more cultural neutral questions. The difficulty levels of the questions should be balance in the questionnaire although Kebbell (2009) argued that questions' difficulties has less influence to the study. In addition, experimenters did not specify whether participants should work alone or not. According to Puncochar and Fox (2004), this may influence the result of this study. In order to ensure that the observed effects were not due to individual differences or practice effects, participants served as their own control for rating of confidence to address the individual differences and they were measured independently to avoid any practice effect. Participants' responses might be made through schemas or heuristics, suggested an area which require attention in future research.

Future research can require participants to justify their responses as it will influence participants' confident level (Koehler, 1991). Koriat, Lichtenstein and Fishchhoff (1980) supported Koehler's (1991) statement as they found overconfidence can be reduced when participants were asked to provide a reason as to why they might be wrong. Future research can vary the risks for being correct or incorrect to motivate the participants as according to Gonzalez-Valejjo and Bonham (2007), both rewarding and punishing consequences produced the best performance. Moreover, researchers can specify whether participants should work alone or not, or maybe include a condition which allows discussion among the participants. Future researchers might be interested to conduct a research using questionnaire comprised of questions from different domains of knowledge. Future experimenters might be interested to examine the role schemas or heuristics into the future research.

This study contributed to the area of criminal justice. From this study, confidence can be a cue towards an eyewitness' statement reliability. Furthermore, current research findings are consistent with findings from previous research from Ordinot, Wolters and van Koppen (2009), Kebbell (2009) and Mengelkamp and Bannert (2010). This strongly supported the results from previous research, which recognize the significant role of eyewitnesses. The finding that time does not affect accuracy can be applied to educational setting, suggesting that students do not need extra time to complete a test, as suggested by Van Overschelde and Nelson (2006) and Koriat, Bjork, Sheffer and Bar (2004), decline in performance is likely to happen when there is more time provided although it is no formally tested in this study. Moreover, this finding supported eyewitnesses when they testify in court as according to current study, time does not affect the accuracy of their statements.

However, more researches can be conducted in this area to draw a strong conclusion. This study can be apply by judges and juries to analyze the reliability of the eyewitnesses' statement, with the suggestion of time is not a good factor to judge the accuracy of eyewitnesses' statement. Law enforcers can judge a statement base on the witness's confidence level as current study found that level of confidence can be a predictor to accuracy, where a statement accompanies with high level of confidence tends to be accurate. In addition, current study suggested that the inaccurate statements which wrongly accused 220 innocent citizens were not influenced by time factor. There are other factors which corrupted the eyewitnesses' statements, which left unexplained by this study. Overall, time does not influence accuracy, confidence and calibration. In making a good judgment, time is not really an important factor due to the existence of schemas and heuristics but it does not mean that good judgment is available for all the time as the reliability of schema and heuristic is questionable. In addition, from this study, time does not influence the reliability of eyewitnesses' statements. This study also supported that confidence level can be used as a predictor or cue towards accuracy of a judgment.

6. References

Ames, D. R., Kammrath, L. K., Suppes, A., & Bolger, N. (2010). Not so fast: The (not-quite-complete) dissociation between accuracy and confidence in thin-slice impressions. *Personality and Social Psychology Bulletin, 36*(2), 264-277.

Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction, 20*(50), 372-382.

Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (1996). Overconfidence in probability and frequency judgments: A critical examination. *Organizational Behaviour and Human Decision Processes, 65*(3), 212-219.

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior, 30*(1), 31-50.

Brewer, W. F., Sampaio, C., & Barlow, M. R. (2004). Confidence and accuracy in the recall of deceptive and nondeceptive sentences. *Journal of Memory and Language, 52*(4), 618-627.

Brewer, W. F., & Sampaio, C. (2006). Processes leading to overconfidence and accuracy in sentences recognition: A metamemory approach. *Memory, 14*(5), 540-552.

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior, 4*(4), 243-260. Doi: 10.1007/BF01040617

Dunning, D., & Perretta, S. (2002). Automaticity and eyewitness accuracy: A 10- to 12- second rule for distinguishing accurate from inaccurate positive identification. *Journal of Applied Psychology, 87*(5), 951-962.

Gonzalez-Valejjo, C., & Bonham, A. (2007). Aligning confidence with accuracy: Revisiting the role of feedback. *Acta Psychologica, 125*(2), 221-239.

Holmes, A. E., & Weaver III, C. A. (2010). Eyewitness memory and misinformation: Are remember/know judgments more reliable than subjective confidence?. *Applied Psychology in Criminal Justice, 6*(1), 47-61.

Kebbell, M. R. (2009). Witness confidence and accuracy: Is a positive relationship maintained for recall under interview conditions?. *Journal of Investigative Psychology and Offender Profiling, 6*, 11-23.

Kebbell, M. R., Evans, L., & Johnson, S. D. (2010). The influence of lawyers' questions on witness accuracy, confidence, and reaction times and on mock jurors' interpretation of witness accuracy. *Journal of Investigative Psychology and Offender Profiling, 7*, 261- 271.

Koehler, D. J. (1991). Explanation, imagination, and confidence in judgement. *Psychological Bulletin, 110*(3), 499-519.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, K. J. (2004). Predicting one's own forgetting: The role of experience-based and theory based processes. *Journal of Experimental Psychology: General, 133*(4), 643-656.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 107-118.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know?. *Organizational Behavior and Human Performance, 20*(2), 159-183.

Lord, R. G. (1985). Accuracy in behavioral measurement: An alternative definition based on raters' cognitive schema and signal detection theory. *Journal of Applied Psychology, 70*(1), 66-71.

Mengelkamp, C., & Bannert, M. (2010). Accuracy of confidence judgments: Stability and generality in the learning process and predictive validity for learning outcome. *Memory and Cognition, 38*(4), 441-451.

Odinot, G., Wolters, G., & van Koppen, P. J. (2009). Eyewitness memory of a supermarket robbery: A case study of accuracy and confidence after 3 months. *Law and Human Behavior, 33*(6), 506-514.

Paese, P. W., & Sniezek, J. A. (1991). Influences on the appropriateness of confidence in judgment: Practice, effort, information, and decision-making. *Organizational Behaviour and Human Decision Processes, 48*(1), 100-131.

Paul, S., & Nazareth, D. L. (2010). Input information complexity, perceived time pressure, and information processing in GSS-based work groups: An experimental investigation using a decision schema to alleviate information overload conditions. *Decision Support Systems, 49*(1), 31-40.

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review, 117*(3), 864-901.

Puncochar, J. M., & Fox, P.W. (2004). Confidence in individual and group decision making: When "two heads" are worse than one. *Journal of Educational Psychology, 96*(30), 582-591.

Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied 15*(1), 46-62.

Sauerland, M., & Sporer, S. L. (2007). Post-decision confidence, decision time, and self- reported decision processes as postdictors of identification accuracy. *Psychology, Crime and Law, 13*(6), 611-625.

Speir-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., & Burgman, M. (2010). Reducing overconfidence in the interval judgments of experts. *Risk Analysis, 30*(3), 512-523.

Van Overschelde, J. P., & Nelson, T. O. (2006). Delayed judgments of learning cause both a decrease in absolute accuracy (calibration) and an increase in relative accuracy. *Memory and Cognition, 34*(7), 1527-1538.

Vickers, D., & Packer, J. (1982). Effects of alternating set for speed of accuracy on response

time, accuracy and confidence in a unidimensional discrimination task. *Acta Psychological, 50*, 179-197. doi: 10.1016/0001-6918%2882%2990006-3

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14*(2), 89-103. Doi: 10.1111/j.1559-1816.1884.tb02223.x

Wright, G., & Ayton, P. (1988). Decision time, subjective probability, and task difficulty. *Memory & Cognition, 16*(2), 176-185. Retrieved from http://ovidsp.tx.ovid.com.ezproxy.lib.monash.edu.au/

Shing, Y. L., Werkle-Bergner, M., Li, S. C., & Linderberger, U. (2009). Committing memory error with high confidence: Older adults do but children don't. *Memory, 17*(2), 169-179.