*Fleisig Dida ***
*Zakay Dan ****

The confidence-frequency effect:
A heuristic process explanation

Abstract: *People's feelings of confidence in the correctness of their knowledge while answering a knowledge test can be inferred in two ways: either by averaging the values of specific confidence values assigned to each item in a test (local confidence) or by asking after the termination of the test for an evaluation of the number of correct answers regarding the entire test (global confidence). Surprisingly, when local and global confidence values of the same test are compared, global confidence tends to be significantly lower than local confidence (the confidence frequency effect). In the present study a heuristic process explanation for the effect is presented and its validity is empirically tested. The global confidence heuristic (GCH) process is based on the ability of participants to recall, after a test was completed, the frequencies of specific confidence values which were assigned to the test's items. Participants build their global confidence by adding about half the number of their guessed answers, to the number of questions with sure answers. The proposed GCH process was supported quantitatively. A content analysis on retrospective explanations provided by participants indicated that this process was feasible. Further research is needed in order to fully explore the power of the explanation suggested here for the confidence-frequency effect.*

Key words: *confidence, confidence-frequency effect, intuition, introspection, local confidence, global confidence*

Introduction

Confidence is a feeling which emerges automatically while accompanying most of goal-directed mental activities (Zakay, 1997). After a mental activity was terminated (i.e., solving a problem, retrieving required information from memory, making a choice, etc.), a retrospective feeling of confidence is being felt. Peterson & Pitz (1988) defined the feeling of retrospective confidence as a belief about the optimality of one's past performance. Retrospective feelings of confidence are important in directing the course of one's future behavior. A domain, in which this is salient, is that of testing people's knowledge by multiple-choice tests. Zakay & Glicksohn (1992) demonstrated how students' feelings of confidence directed their behavior in such tests. Ongoing feelings of confidence may determine whether a specific planned task will be executed and whether or not an ongoing activity will be stopped or continued(Zakay, 1997). Kruglanski, Peri and Zakay (1991) showed that when the initial level of confidence in a given hypothesis was high, the motivation for acquiring more information in order to test the hypothesis was lower than when the initial confidence was low.

People's confidence in the correctness of their answers in a test is traditionally measured in two ways: probabilistic confidence judgments relating to each question ("specific confidence judgments") and frequency judgments relating to the total of the test's questions ("global confidence judgments"). Specific confidence is measured by asking a participant to assign each one of his/her responses a value on a 50-100 probability scale (in the case of a two alternatives questions). By assigning the value of "100" the participant indicates that he/she is sure about the correctness of the response. By assigning a value of "50", the participant indicates that the response was based on a guess. The average of all specific confidence values is considered to reflect the level of confidence regarding the entire test.

* **College of Management Academic studies, Rishon Lezion, Israel, didafl@gmail.com**
** **Interdisciplinary Center (IDC), Herzlyia, Israel**

Global confidence is measured by asking a participant to evaluate the number of correct responses in the whole test. Thus, local and global confidence should both indicate the overall level of confidence in the entire test. Surprisingly, it was found that when local and global confidence regarding the same test were compared, a paradoxical state was revealed: Local confidence was mostly higher than global confidence. According to Hoffrage (2004), local confidence exceeded global confidence by about 15 percent. This discrepancy, which was termed "the confidence-frequency effect" by Gigerenzer, Hoffrage, & Kleinbolting, (1991); or the "aggregation effect" by Treadwell & Nelson, (1996) was demonstrated in numerous studies regarding general-knowledge tests. (May, 1986; Griffin & Tversky, 1992; Keren, 1991; Sniezek, Paese, & Switzer, 1990; Schneider, 1995), as well as in evaluating the quality of physical performance in skill tasks (Stone, Rittmayer, Murray, & Murray McNiel, 2011).

Several explanations were provided for the "confidence-frequency effect": Schneider (1995) as well as Sniezek et al. (1990) proposed the Dual-Source Explanation according to which each of the confidence judgments relied on information concerning distinctly different issues, each one relied on information coming from a different source. Therefore these judgments should not be expected to yield similar results. This explanation is not accounting for the effect since it is vague and too general.

Keren (1991) argued that as specific confidence was restricted to a 50-100 scale (In double-choice questions) and the global confidence scale ranged between 0-100, a discrepancy between both judgments was not inevitable. Indeed, when below-chance judgments were excluded from the analyses, no difference between local and global confidence was found. However, that explanation was not supported in all experiments (e.g., Sniezek, Paese & Switzer, 1990)

Liberman (2004) suggested that the confidence-frequency effect stemmed from participants' failure to understand that when guessing (In binary tests) they still had a 50% chance of being correct. In one experiment, participants answered double-choice questions, and scored their confidence in the correctness of each answer (specific confidence). Global confidence was judged differently in each experimental group: participants in the "unrestricted group" answered the question: "what percentage of questions have you answered correctly?" while participants in the "restricted group" were additionally told that "Your estimate should be 50% or higher, because if you were to answer at random about 50% of the answers would be correct" (Liberman, 2004, p. 729). It was found that global confidence was lower than local confidence in the "unrestricted" group, but was higher than local confidence in the "restricted" group.

Griffin & Tversky (1992) proposed a resembling explanation as they found that participants reported as correct (global confidence) only the items they knew with certainty, and did not include any items they guessed. Some participants added an alternative explanation that proposed that unlike specific confidence, the global confidence

reflected the difficulty of the test. Griffin & Tversky as well asLiberman's explanations are partial, since they refer to one category of responses only.

Griffin & Buehler (1999) conducted an experiment, in which participants were asked to answer 20 general knowledge double-choice questions and rated their specific and global confidence. The researchers additionally asked the participants to explain in writing how they evaluated their global judgment. The confidence-frequency effect was revealed, namely, participants were overconfident in their local confidence, and underconfident in their global confidence. Analyses of the verbal explanations provided by the participants enabled the identification of three strategies by which global confidence judgments were obtained. Two strategies were based on specific confidence judgments. According to one strategy, participants rated their global confidence as the number of questions for which they assigned the specific confidence value of 100, while ignoring all other answers. The size of the confidence-frequency effect for this group was the biggest. According to a second strategy participants counted the number of answers for which they assigned fairly confident values and to this they added a number of questions to account for the answers they guessed. The size of the confidence-frequency effect in this group was smaller than that found in the former group. A third strategy, which was not based on specific confidence values was to evaluate global confidence on the basis of cues like the felt difficulty of the test, the participants' belief in their knowledge and their experience in answering similar tests. The size of the confidence-frequency effect in this group was minimal.

The global-confidence heuristic (GCH) process

Based on Griffin & Tversky (1992), Griffin & Buehler (1999), and Liberman (2004), and in accordance with Gigerenzer et al. (1991) who suggested that specific-confidence ratings could serve as relevant cues to global confidence ratings, we propose that when participants are called to perform global confidence judgments regarding their performance in a test, they use a heuristic process which we call the global-confidence heuristic (GCH) (Fleisig, 2005; Zakay & Fleisig, 2009; 2011). The proposed GCH process is composed of the following steps: participants retrieve the frequencies of responses according to categories of specific confidence values which were assigned during a test. The number of responses which were assigned specific confidence values of 100 (sure answers) are stored in memory. To this value, about half of the number of responses to which specific confidence values of 50 were assigned (guesses) is added. The obtained value is considered to reflect the requested global confidence. For example, suppose that a participant was presented with a forced choice test, composed of 10 binary questions. The participant assigned specific confidence values of 100 to three responses and specific confidence values of 50 to other four responses. Our participant assigned to the remaining three responses specific confidence values of 70. When requested to evaluate her global confidence (the overall number of correct responses), the participant counted as

correct al three sure responses (100) and added two of the four guessed ones, while ignoring the rest, thus reaching the value of 5. (Note that the value of local confidence which is the average of the specific confidence values is, in this case, 7.1). This heuristic process reflects a naïve theory according to which all answers which were assigned a specific confidence value of 100 are correct and about 50% of the answers which were guessed should also be correct. Liberman (2004) found that not all participants considered guessed questions. In such a case, and in accordance with the first strategy identified by Griffin & Buehler (1999), the GCH process would be comprised only of the first step described above, namely, counting the number of sure answers. It is clear that the GCH process is not imposing high mental load on participants, since it does not require accurate calculations.

The explanation of the confidence-frequency effect, according to the Global-Confidence heuristic process is that the global confidence judgment reflects under-representation of answers that were not assigned with high confidence scores. When subjects count as correct all the "certain" answers but do not allocate a high enough probability for the correctness of the other answers, about which they felt less confident, and thus do not count some of them as being correct, global judgment is, eventually, lower than the average of the specific confidence judgments, i.e. local confidence, hence resulting in the confidence-frequency effect.

The aim of the present study is to validate the proposed heuristic process. In order to do so, the following three hypotheses should be empirically supported: 1). The confidence-frequency effect will be found. 2). Participants can retrieve their specific confidence judgments and the frequencies of the judgments after a knowledge test was completed. 3). Participants will describe the process by which they evaluated global confidence values and this process will reflect the GCH process. It should be noted that validating the second hypothesis is crucial, and this is directly tested for the first time in the presents study.

Empirical testing of the GCH process

The proposed heuristic process describes the way participants evaluate their global confidence – by way of relying on their specific confidence judgments.

The purpose of the experiment was to validate the GCH process by prooving participants' ability to recall the frequencies of their specific confidence judgments, and to demonstrate their use of the proposed heuristic process.

Method

Participants:

40 students (7 males and 33 females) participated in the experiment, as partial fulfillment of course requirements. Mean age was 22.6 years (ranging between 18 to 30 years; sd=2.27). All participants were naïve as to the purpose of the experiment.

Materials:

General knowledge test: a general knowledge test comprised of 40 binary items was constructed. (Examples of questions asked are: a Spanish island in the Mediterranean sea is: 1. Ibiza or 2. Bermuda.; The author of "The Satanic Verses" is 1. Salman Rushdie or 2. Nagib Mahfuz.). No misleading questions in terms of a confounding between familiarity and accuracy (for example: the capital of Switzerland is Bern, but Zurich is much more familiar) were included. All questions had a correct answer and the questions represented topics which are taught in Israeli high schools (e.g., history, geography, etc.).

Procedure: The experiment was conducted with an IBM Pentium PC. Following Allwood, and Montgomery (1987). The subsequent instructions appeared on the computer screen: "*you will be presented with questions having two alternative answers, marked by the numbers 1 and 2. You are asked to choose the correct answer by entering the number of that answer on the keyboard. After you have answered, you will be asked to rate the degree of your confidence in the accuracy of your answer on a scale that ranges between 50 and 100. A score of 50 expresses a guess and a score of 100 expresses complete confidence in the correctness of your answer. The scale ranges between 50 and 100 because if you choose a score which is lower than 50 to mark a certain answer, it means that you believe the second answer should get a score higher than 50. If this is the case, then you were supposed to choose that answer as the correct one. You may use all values between 50 and 100, with higher values expressing higher confidence*".

After a participant completed answering all 40 questions, the following instructions appeared on the screen: "*How many questions, out of the 40 you were presented with, did you, in your opinion, answer correctly?*"the participant entered any number between 0-40, which reflected his/her global confidence.

In the next step, participants were asked to write on a paper an explanation describing the way they evaluated their global confidence. They were asked to be as accurate and detailed as possible.

After completing the written explanation, the participant was presented with a table consisting of 2 columns and 10 rows. The heading on top of the first (on the right) column was: "*confidence score*" and the heading on the second column (on the left) was "*number of questions*". The instructions were: "*in the following table you are asked to write the confidence scores you assigned all of the 40 questions in the test. For each "confidence score" (i.e. 50, 63, etc.) you should indicate the number of questions that, according to your memory, were assigned that score. For example, if you remember assigning a confidence score of 72 to three questions, you should write "72" on the right column and "3" on the left*". Afterwards, the following instructions appeared on the computer screen: "*you will be presented with categories of confidence scores. Please state the number of questions to which you assigned a confidence score within that category*". The specific confidence categories appeared one at a time on the computer screen, and

participants entered the number they remembered to be the number of questions they assigned those confidence scores in that category. The following seven confidence categories (based on Gigerenzer et. Al, 1991) were presented: 50; 51-59; 60-69; 70-79; 80-89; 90-99; and finally category 100.

To summarize, the structure of the experiment was as follows:
1. General knowledge test including assigning specific confidence values
2. Global confidence judgment
3. Written explanation of the judgment of global confidence
4. Free-recall of frequencies of assigned specific confidence values
5. Structured recall of frequencies of assigned specific confidence values, regarding each of 7 presented categories

Participants performed the experiment individually. The test was self paced, and mean completion time was 30 minutes.

Results

Hypothesis 1: The confidence-frequency effect

Local confidence values were calculated for each participant and compared to respective global confidence values. The confidence-frequency effect was revealed, as average local confidence was higher than average global confidence (M=76.88, sd=6.99 and M=57.25, sd=19.67, respectively; T-test for dependent variables T=-7.68, p<.0000). Participants were overconfident in their local judgment, as it was higher than actual performance (M=70.94, sd=10.97 in percentage ; T-test for dependent variables T=-3.69, p<.001). Participants were under confident in their global confidence. (T-test for dependent variables T=4.90, p<.001).

Hypothesis 2: Recollection of specific confidence scores

Participants were asked to recall the frequencies of their specific confidence scores regarding their answers in the test in two ways: "free recollection" (filling the table with confidence scores and their frequencies – stage 4) and by "structured recollection" (indicating frequencies to presented confidence categories – stage 5). Based on those frequencies, two values of recollected local confidence were calculated for each participant: the mean of the "free" recalled scores (F-R), and the mean of the "structured" recalled scores (S-R). Means and SD (in brackets) of the actual local confidence(A-LC), and the recalled confidence scores (F-R, and S-R) were the following: 76.88 (6.99); 75.67 (7.79); 77.47 (7.19); respectively.

The three confidence scores were compared by ANOVA with repeated measures. No significant differences were found (F(2, 74)=.71, n.s.).

Pearson Correlations between the values of F-R and A-LC and between S-R and A-LC were computed for each one of the seven confidence categories. The correlations are presented in table 1.

Table 1: Pearson correlations between F-R, S-R, and A-LC values

Confidence category	100	90-99	80-89	70-79	60-69	51-59	50
Correlation F-R and A-LC	.43*	.33*	.52*	.17	.38*	.54*	.48*
Correlation S-R and A-LC	.76*	.48*	.51*	.14	.42*	.47*	.68*

Note: p<.05 *

The results support hypothesis 2. Participants were able to recall the frequencies of the specific confidence values they assigned to the test items, after the test was completed. Local confidence values computed by averaging the recalled and actual specific confidence values did not differ significantly from each other. It is of interest to note that the high confidence categories (90-99; 100) and the "guess" or "near guess" categories (50; 51-59) were most accurately recalled.

Analysis of the retrospective explanations

The analysis of the retrospective explanations was done in two stages. In the first stage, two expert judges extracted and defined process units reflecting basic steps in the process, for each one of the retrospective explanations. After comparing the product of each one of the experts, 9 common process units were defined. In the second stage, the experts reanalyzed the retrospective explanations and identified those process units which appeared in each retrospective explanation. In order to permit the completeness of the analysis, 3 more units which were defined in previous research (Fleisig, 2005) were added, thus amounting the total to 12 units. The compatibility between the two experts was very high (above .90).

A note on process tracing

Part of the analyses in the experiment relies on participants' retrospective verbal explanations to the way they evaluated their global confidence values. The notion of extracting thinking processes from verbal protocols goes back to Newell and Simon (1961). In order to gain understanding of human decision processes, one can not study final decisions alone, but has to examine additionally the perceptual, emotional, and cognitive processes (Svenson, 1979; 2003; Ranyard & Svenson, 2011).Process tracing can help in assessing what information was used in forming a judgment (Payne, Braunstein, and CarrolI, 1978). Verbal protocols (amongst others) enable inferring the decision strategies that were utilized by participants. (Ford, Schmitt, Schechtman, Hults, & Doherty, 1989).

As Schulte-Mecklenbeck, Kühberger, and Ranyard (2011) state (p. 733): "Process models can be tested and

evaluated in terms of both JDM behavior and process tracing methods, which elicit and analyze observations of a range of verbal and nonverbal antecedents and concomitants of judgments and decisions." The authors argue that for process models validation process data is important, because it is richer than input-output data and can provide important evidence of explanatory mechanisms. In the present study, the explanation for the confidence-frequency effect is based on a heuristic process. Relying on process tracing methods in order to validate the suggested heuristic process seems relevant and justified.

Results of the content analysis of the retrospective explanations

The process units which were extracted from the retrospective explanations and from previous research are represented below. Alongside each process unit a verbatim from the written explanations is presented in brackets.

1. Referring to feelings of confidence ("*I counted the questions according to the feelings I had, whether I was confident in my answer or not*")
2. Counting the questions for which a specific confidence value of 100 was granted, and "correcting" that number ("*I think that I scored 100 in confidence regarding about 4 questions, but I might be wrong, so I'll count 3 as certainly correct*")
3. Counting the questions for which a specific confidence value of 100 was granted, yet, not "correcting" that number ("*I scored 100 in confidence for 5 questions, so all 5 are certainly correct*")
4. Counting the questions for which a specific confidence value of 50 was granted and "correcting" that number ("*I guessed 6 questions, I really had no clue, yet, as there were only 2 answers, I have a chance that half of them are correct, so out of the 6 I'll count 3 as correct*")
5. Counting the questions for which intermediate specific confidence values were granted, and "correcting" that number ("*I think that regarding about 4 questions I was not sure, I did not guess but something in between, so I have the chance that 3 will be correct*")
6. A manifestation of a probabilistic thinking regarding the specific confidence values, and "correcting" the calculated number ("*I assume I should calculate the sure answers, the guesses and the others according to the probability of them being correct, because even when I am certain I still might be wrong, and maybe my guess was correct?*")
7. Counting questions based on specific confidence values without "correcting" that number ("*I was not so sure about 5 questions, but I'll assume they still are correct*")
8. A manifestation of a probabilistic thinking not regarding the specific confidence values("*I think I have the chance of half of the answers being correct*")
9. A reference to the overall difficulty of the test ("*The test was difficult, so I might have succeeded in 4*")
10. A reference to a personal capability ("*I am really bad in geography, so let's say 3*")

11. Pure guess ("*I just guessed – 7*")
12. Restoring the actual answers ("*there was that question about Belgium which I think I answered Ostend, and that question about Tanzania....*")

It should be noted that whereas process units 1-7 refer to confidence feelings, process units 8-12 are not associated with specific confidence.

The frequencies of appearance of each process unit across all 40 retrospective explanations are presented in table 2.

Table 2:frequencies of appearance of each process unit across all 40 retrospective explanations.

Process unit	frequency
1- referring to confidence	7
2 - counting scores of "100" and correcting	0*
3 - counting scores of "100" without correcting	15
4 - counting scores of "50" and correcting	14
5 - counting intermediate scores and correcting	2
6 - probabilistic thinking regarding the confidence scores	9
7 - counting questions based on confidence scores without correcting	4
- Total for categories regarding specific confidence	51
8 - probabilistic thinking not regarding confidence scores	7
9 - reference to the overall difficulty of the test	0*
10 - reference to a personal capability	4
11 - pure guess	2
12 - restoring the actual answers	0*
- Total for categories not regarding specific confidence	13

Note: * added from previous research (Fleisig, 2005)

As can be seen, most explanations (51) were based on specific confidence, whereas only 13 were not. Out of the explanations that were based on specific confidence judgments a pattern emerged: counting the number of questions for which the specific value "100" was granted, as correct answers, and counting the questions for which the specific value "50" was granted and adjusting that number for the global assessment. However, it should be noted that different participants utilized different strategies. It should also be noted that the frequencies of the added process units in the second stage analysis was zero.

Comparing global confidence values based on the content analysis

Participants were divided into two groups, based on the nature of their retrospective explanations. The "confidence" group was comprised of 29 participants who based their global confidence evaluations on specific confidence, in one way or another. The "intuitive-confidence" group was comprised of 11 participants who did not base their global confidence evaluations on specific confidence.

No difference (In a one-way Anova, N.S.) was found between the two groups regarding their actual test scores (*M*=71.29, *sd*=12.06 for the "confidence" group; *M*=70.00 *sd*=7.82 for the "intuitive-confidence" group). Similarly, no difference (In a one-way Anova, N.S.) was found in their local confidence values (*M*=75.98, *sd*=7.25 for the "confidence" group; *M*=79.26 *sd*=8.13 for the "intuitive-confidence" group). Yet, a significant difference (in a one-way Anova $F(1, 38)=5.70$, *p*=.022)was found between the two groups in their global confidence values (*M*=52.93, *sd*=19.92 in the "confidence" group; *M*=68.63 *sd*=14.15 in the "intuitive-confidence" group) . It emerges that the confidence-frequency effect is significantly higher in the "confidence" group (*M*=23) than in the "intuitive-confidence" group (*M*=11) ($F(1,38)=5.22$, *p*<.02).

Discussion

The picture that emerges from the joint quantitative and qualitative analyses is the following:

The confidence-frequency-effect was obtained, however, its size was found to be dependent on the type of process participants utilized in order to evaluate their global confidence. The effect was large for those participants who were using the complete CGH process (basing their evaluation on both "sure" and the "guesses" categories) or partially (basing their evaluation on the "sure" category only).

The confidence-frequency-effect was also revealed for participants who were not relying on specific confidence values (intuitive-confidence group) while evaluating their global confidence, but it was much lower than in the former group (23.05 and 10.63 respectively).

It seems that people use different strategies when facing the need to evaluate global confidence. Most people (72.5% in our sample) tend to rely in some way on the CGH process, and fewer (27.5% in our sample) are relying on intuitive feelings about the test difficulty and their past experiences in similar tests. It was interesting to find that despite the fact that the instructions favored numerical estimates as indicators of global confidence, about one fourth of the participants used another heuristic.

It is also of interest to note that relying on an analytical process like the CGH , is leading to a less accurate evaluation of global confidence than when relying on a pure intuitive process. The reason is that while applying the CGH process, people used only the most salient (i.e., 100 and 50) categories of specific confidence values. This might serve as an illustration for a situation in which a utilization of a system 2 process, which demands resources (Kahneman, 2011) is less optimal than using pure intuition (system 1).

Overall, we have validated in this study the ability of people to recall the frequencies of specific confidence values, and to use the CGH process in full, or partially. However, not everyone is doing so, and some are relying on "guts-feelings". This finding is of interest since, as was discussed earlier, feelings of confidence might determine one's future behavior. Thus, being accurate in evaluation one's confidence is contributing to behavior's optimality.

In the present paper it was demonstrated that confidence's accuracy is dependent on the type of process used for evaluating level of confidence. A potential implication of this is that it might be possible to educate people how to increase their accuracy regarding their confidence. Further research is needed in order to fully explore this idea.

References

Allwood, C. M., & Montgomery, H. (1987). Response selection strategies and realism of confidence judgments. *Organizational Behavior and Human Decision Processes, 39*, 365-383.

Fleisig, D. (2005). *The confidence weighing (CW) heuristic: An explanation to the overconfidence paradox*. Ph.D. Dissertation (in Hebrew). Tel Aviv University.

Ford, J.K., Schmitt, N., Schechtman, S.L., Hults, B.M., & Doherty, M.L. (1989). Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior and Human Decision Processes, 43(1),* 75–117.

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506–528.

Griffin, D., & Buehler, R. (1999). Frequency, probability, and prediction: Easy solutions to cognitive illusions? *Cognitive Psychology, 38*, 48–78.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411–435.

Hoffrage, U. (2004). Overconfidence. In R.F. Pohl (ED). *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgments and Memory.* Psychology Press.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux, New York.

Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica, 77*, 217–273.

Kruglanski, A.W., Peri, N., and Zakay, D. (1991). Interactive effects of need for closure and initial confidence on social information seeking. *Social Cognition, 9,* 127-148.

Liberman, V. (2004). Local and global judgments of confidence. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 729–732.

May, R. M. (1986). Inferences, subjective probability and frequency of correct answers: A cognitive approach to the overconfidence phenomenon. In: B. Brehmer., H. Jungerman., P. Laurens., and A. Sevon (Eds.), *New directions in research on decision making* (pp. 175–189). Amsterdam: North-Holland.

Moore, D.A., & Healy, P.J. (2008). The trouble with overconfidence. *Psychological Review, 115 (2)*, 502-517.

Newell, A., & Simon, H. A. (1961). Computer simulation of human thinking. *Science, 134*, 2011-2017.

Payne, J. W., Braunstein, M. L., & CarrolI, J. S. (1978). Exploring pre-decisional behavior: An alternative approach to decision research. *Organizational Behavior and Human Performance, 22*, 17-44.

Peterson, D. K., & Pitz, G. F. (1988). Confidence, uncertainty, and the use of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 85–92.

Ranyard, R, & Svenson, O. (2011). Verbal data and decision process analysis. In M. Schulte-Mecklenbeck, A. Kühberger, and R.Ranyard, (Eds.), *A handbook of process tracing methods for decision research: A critical review and user's guide.* Society for Judgment and Decision Making series., (pp. 115-137). New York, NY, US: Psychology Press, xv, 256 pp.

Schneider, S. L. (1995). Item difficulty, discrimination, and the confidence-frequency effect in a categorical judgment task. *Organizational Behavior and Human Decision Processes, 61,* 148–167.

Schulte-Mecklenbeck, M., Kühberger, A & Ranyard, R. (2011). The role of process data in the development and testing of process models of judgment and decision making. *Judgment and Decision Making, 6(8),* 733-739.

Sniezek, J. A., Paese, P. W., & Switzer, F. S. (1990). The effect of choosing on confidence in choice. *Organizational Behavior and Human Decision Processes, 46*, 264–282.

Stone, E.R., Rittmayer, A.D. , Murray, N.T., & Murray McNiel, J. (2011).

Demonstrating and eliminating the aggregation effect in physical skill tasks. *Journal of Behavioral Decision Making, 24 (1)*, 1–22.

Svenson, O. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance, 23*, 86-112.

Svenson, O. (2003). Values, affect and processes in human decision making: a differentiation and consolidation theory perspective. In S.L. Schneider & J. Shanteau (Eds.) *Emerging Perspectives on Judgment and Decision Research.* Cambridge: Cambridge University Press, (pp. 287-326).

Treadwell, J. R., & Nelson, T. O. (1996). Availability of information and the aggregation of confidence in prior decisions. *Organizational Behavior and Human Decision Processes, 68(1)*, 13-27.

Zakay, D. (1997). Post-decisional confidence – can it be trusted? In: R. Ranyard, R. Crozier, and O. Svenson (Eds.), *Decision making: cognitive models and explanations* (pp. 233–246). London: Routledge.

Zakay, D., & Fleisig, D. (2009). The overconfidence paradox in perceptual tasks. In M. Elliot (Ed.), *Fechner Day 2009* (pp. 266–280). Galway: National University of Ireland.

Zakay,D., & Fleisig, D. (2011). Confidence-Frequency discrepancy in judgments. In W. Brun, G. Keren, G., Kirkeboen & H. Montgomery (Eds.) *Perspectives on Thinking, Judging and Decision Making.* Springer, Universitetsforlaget.

Zakay, D. & Glicksohn, J. (1992). Overconfidence in a multiple-choice test and its relationship with achievement. *Psychological Record, 42*, 519-524.