

COMPARING RESPONSE LATENCY AND SELF-REPORT METHODS FOR ESTABLISHING LEVELS OF CERTAINTY EMPLOYING THE AHP

Jerald Feinstein, Erik Winslow, Thomas Nagy, and John Coyne
The George Washington University
School of Business and Public Management
Washington, DC, USA
jerry@gwu.edu

Abstract

This investigation is concerned with using the Analytic Hierarchy Process [AHP] to compare two methods for estimating a subject's feeling of certainty or confidence in their choice. One method is the older or classical method where a subject self-reports both the choice selected as well as the level of certainty associated with a choice-pair on the one-to-nine scale used with the AHP. The second is proposed by the authors and based on using only the choice and decision time or response latency. In the second case, the inverse of response latency is used to estimate a subject's degree of certainty or confidence in the selected choice in a paired comparison experiment. The response latency method has the advantage of being unobtrusive, less prone to conscious censure, quicker to perform, requires less effort, and possibly is less expensive to administer.

Employing AHP methodology, priorities are used to represent degree of confidence among choices, and consistency ratios are used to estimate the degree of consistency detected by each method. To compare the two methods, 21 subjects were evaluated in a paired sample design. Employing a computer-administered questionnaire, subjects expressed relative levels of confidence for a set of potential recommendations and the time for each subject to respond was recorded unobtrusively.

Using paired sample t-tests, sufficient evidence was found to reject the null hypotheses ($\alpha = .025$) that the mean of the difference in the priorities between the two methods was zero ($p < .0001$). In a second test, there was sufficient evidence to be at least 95% confident of accepting an alternative hypothesis that response latency detects a lower consistency ratio than the self-report method ($p < .0001$). In addition to the tests of significance, the results were found to be practically significant and are discussed in the paper.

LITERATURE REVIEW

Currently, the procedure employed to estimate degree of certainty is to ask subjects to self-report their degree of preference, confidence, or certainty that they associate with a particular recommendation [Harmon and King, 1985]. Employing the self-report approach, subjects must report not only their preferred choice, they must provide an estimate of their confidence for their choice compared to other possible ones. Reporting the preferred choice is a relatively easy task for the subject. This is because people are accustomed to making choices. However, quantifying their degree of confidence in one choice over others is something that people find difficult, as they are being asked to quantify something that they do not normally quantify in practice. To further complicate the situation, conscious processes are involved when deciding on relative levels of confidence and these processes are known to be vulnerable to manipulation. Subjects sometimes respond to questions in a manner that they think pleases the questioner, or respond in ways that tend to maintain their image of themselves. Thus, the self-report approach is not only time-consuming, it is known to be vulnerable to conscious censure and manipulation [Clemen, 1996].

A subject's self-reported preferences or estimates of certainty are often inconsistent and may not be a good measure of their true attitude. For example, subjects may be consciously unaware of their true feelings [Banaji and Greenwald, 1993; Greenwald & Banaji, 1995; Nisbett and Wilson, 1977], or they may be reluctant to reveal their true feelings [Crosby, Bromley, and Saxe, 1980; Gaertner and Dovidio, 1986; Sigall and Page, 1971].

Self-reported information may be subject to further degradation as subjects can become annoyed with the time-consuming and often uncomfortable process of not only responding to which is the preferred recommendation, but being asked to estimate how much the decision selected is preferred to other ones [Marshall and Oliver, 1995]. Subjects often develop anxiety over the requirement of directly reporting a degree of preference [Medsker, 1998]. If subjects are uncomfortable with the feeling that they might provide an incorrect answer, their certainty may waver [Medsker, 1998], or they may view the process as a waste of time [Marshall and Oliver, 1995] and provide inaccurate responses. Such concerns about the criterion validity of self-reported information underscore the problem and the need for using more unobtrusive measures [Dovidio and Fazio, 1992].

Response latency may provide a solution to the self-reporting problems because it is believed that the faster a choice is made, the stronger the subject's relative certainty of the choice or degree of preference; and the approach is well founded. A substantial body of literature supports response latency as a measure of preference and certainty in psychological research and market analysis studies [LaBarbera & MacLachlan, 1979]. However, its methodological use in estimating levels of certainty and consistency in decisions appears almost nonexistent.

Earlier work by Dashiell [1937], Cartwright [1941], Festinger [1943], and Tyebjee [1979] provide evidence that response latency tends to be inversely related to the degree of certainty experienced in making a judgment and inversely related to the objective or subjective distances between the choice-pairs; that response latency could be used to estimate the difference in affective value across multiple decision variables, and is insensitive to the presentation order of the choice-pairs. Yet, there appears to be no well-defined or established methodology where response latency is used to estimate a subject's level of confidence or preference.

The AHP introduced by Saaty [1977] provides a methodology for generating a ratio scale of a subject's relative levels of certainty or preference over a set of choices using pairwise comparisons. The certainty values are called priorities and are derived from the elements of an eigenvector formed from the matrix of pairwise comparisons. Also provided is a measure of consistency called a Consistency Ratio [CR] defined by the following formula,

$$CR = \lambda_A - \text{order [A]} / [\text{order [A]} - 1] \lambda_{\text{random}}$$

where λ_A is the largest eigenvalue of the reciprocal matrix of pairwise comparisons A, and λ_{random} is the largest eigenvalue of a randomly generated reciprocal matrix of the same order as matrix A, and Order [A] is the order of matrix A.

Saaty [1977] defines a one-to-nine scale for self-reporting degrees of certainty or preference. A "one" signifies equal confidence or preference between the choice-pairs presented and a "nine" indicates an order of magnitude difference. In assessing degree of certainty, the subject would be asked which choice is preferred [higher confidence] and how much that choice is preferred to the other.

Thus, the AHP is an excellent methodology to compare the response latency and self-report approaches. In the self-report case, the self-report data from the one-to-nine scale would be used as input to the matrix of pairwise comparisons. For the response latency case, the inverse of response latency would be used where response latency is the time taken to make a choice in a pairwise comparison of choices.

PROCEDURE

Using a Pentium computer and benchmarked psychological testing software, Micro-Electronic Laboratory [MEL2] version 2.01, from Psychological Software Tools of Pittsburgh, 21 subjects were asked to express their preference for one of two investment choices presented in random order on a computer monitor as well as the level of confidence in their selection. Response latency was collected unobtrusively for choice selection as well as the length of time for each subject to read the instructions.

The choices were [1] Cash In Hand, [2] Asian Growth Fund, [3] S&P Index Fund, [4] Blue Chip Fund, and [5] Money Market Fund. Each subject was asked to complete a set of questions from one block of comparisons and then another. The two blocks were presented to each subject in random order and

represented two different economic scenarios each describing a different set of rule conditions. One scenario represented a world-wide economic boom and the other a world-wide economic catastrophe or bust.

Subjects were required to use only two keys on the keyboard to signify which choice was preferred; "Q" [color-coded red] was used if they felt that the choice on the left side of the screen was preferred, or "P" [color-coded yellow] if the choice on the right side of the screen was preferred. After the subjects selected the preferred choice, they were asked to input a level of confidence by entering a number between one and nine, representing a range from equal confidence to almost perfect confidence in the choice selected over the other.

For each subject, the information recorded by the computer was: The choice-pair offered, the preferred choice, the subject's estimate of certainty, and the response latency or time it took for the subject to identify the correct choice. The last item is being collected unobtrusively and without the knowledge of the subject. In order to learn if subjects were reading the instructions, individual reading times were recorded. However, subjects reported that they became familiar with the instructions during the briefing and practice period to the degree that some felt that they only needed the first few lines of the instructions before continuing. In these cases, the time spent reading the instructions was relatively brief.

EXPERIMENTAL DESIGN

A crossover design and paired sample *t*-test is used to test of significance for the hypotheses. The rationale for the paired sample *t*-test is that the statistical assumptions for the standard *t*-test are not satisfied as the subject's level of certainty is measured twice - - once for each method. Therefore, the samples are not independent, as required by the *t*-test. The advantage of this approach is that it eliminates the variation between subjects. Therefore one need not be concerned about variations in the subjects' natural speed in making decisions due to age or other factors, reaction times, variation in reading speed, or any other factors that could affect the subjects differently.

The approach selected to choose between the two techniques should each produce different sets of Consistency Estimates was to assess the degree to which each method detects each subject's consistency across the choice-pairs. If there is some underlying consistency in the subject's set of Certainty Estimates, then the preferred method would be the one that detects those consistencies better than others. It should be emphasized that consistency is neither good nor bad. Often, many highly accurate predictions are inconsistent. One example is in team sports where Team A can beat Team B, Team B can Beat Team C; however, Team C always seems to beat Team A. For perfect consistency, Team A would beat Team C. Yet, whatever the level of consistency present, the method that detects consistency better than the other is the one preferred. Thus, if we view the same process through two different instruments, the instrument defined as better is the one possessing the lower Consistency Ratio, as it detects consistency better than the other.

SUMMARY OF APPROACH

Data from Subject 5 are used as an example to illustrate the process of calculating Certainty Estimates and a consistency ratio for one subject. For each choice-pair the subject responds with the preferred choice. After the subject responds, both the choice and the time to respond are recorded by the computer. Then, on a scale of one-to-nine, subjects enter their relative degree of confidence in the choice selected to the other. A "one" signifies equal confidence as in "flipping a coin to decide," and a "nine" represents very high confidence, as in "almost a sure thing." This self-reported feeling of confidence also is recorded. Response latency is measured by the computer in increments of hundredths of a second. If the latency is greater than nine seconds, a value of nine seconds is used in the analysis. This level of truncation is of the same order of magnitude used by other researchers [LaBarbera & MacLachlan, 1979].

Next, the inverse of the response latency for a choice is calculated with dimensions of 1/dekaseconds, and is used as the measure of relative certainty. Then, as prescribed in the AHP, if the column variable dominates the row variable within the matrix of pairwise comparisons, the inverse of response latency; if the row variable dominates the column variable, then the response latency [in dekaseconds] is used. Dominance is determined by the choice selected from the choice-pair presented.

The procedure is illustrated by employing actual data collected from Subject 5 as shown in Figure 4. Matrix of Pairwise Comparisons, Subject 5, for the Response Latency Method. The rows and columns are the investment choices offered: CIH – Cash In Hand, AGF – Asian Growth Fund, SPI – S&P Index Fund, BCF – Blue Chip Fund, and MMF – U.S. Government Money Market Fund. Each element, a_{ij} , represents the subject's confidence in selecting choice i over j . If "i" is preferred to "j" then the rescaled value is used, else the response latency in dekaseconds is used. If both are preferred equally, the value used is 1. The elements below the diagonal are the inverses of corresponding values above the diagonal [$a_{ji}=1/a_{ij}$]. In this example, AGF is chosen over CIH, and $a_{12}=.206$, the value for the response latency in dekaseconds. That signifies that the subject was about 1/5 as certain about CIH than about AGF. However, SPI is chosen over MMF and $a_{35}=4.37$, the inverse of the response latency in dekaseconds. This means that the subject was more than four times as certain about SPI than about MMF.

First, CIH is compared with each variable along the row. First itself, then, AGF, next, SPI to MMF. If CIH is selected over the variable it is being compared to, then the inverse is used, if they are the same, a one is used, and if the other variable is selected, the response latency in dekaseconds is used. Then, AGF is compared to itself, then SPI, to MMF. Each element above the diagonal is entered in this manner, and the elements below the diagonal are entered as reciprocals as indicated above. The resulting matrix is

illustrated as Figure 1. Matrix of Pairwise Comparisons, Subject 5, for the Response Latency Method. The results are illustrated exactly as output from Excel. For the self-report method, the same technique is used to enter values into the matrix. However, in this case, the raw data is already within the one-to-nine range, so the values can be entered directly. The AHP dominance criterion is used to enter the actual value or its reciprocal. The resulting matrix of values for Subject 5 is illustrated as Figure 2. Matrix of Pairwise Comparisons, Subject 5, for the Self-Report Method.

For Subject 5, the eigenvalue method is used to calculate a set of priorities from the response latency data and another set of priorities from the self-reported data. In addition to the two sets of priorities, a Consistency Ratio is calculated for each set of priorities.

	CIH	AGF	SPI	BCF	MMF
CIH	1	0.166666667	0.142857143	0.142857143	0.142857143
AGF	6	1	0.142857143	0.142857143	0.142857143
SPI	7	7	1	0.142857143	7
BCF	7	7	7	1	7
MMF	7	7	0.142857143	0.142857143	1

Figure 1. Matrix of Pairwise Comparisons, Subject 5, for the Self-Report Method

To illustrate the approach, Subject 5's Certainty Estimates [shown as priorities] for the different choices as well as the associated Consistency Ratio are illustrated for a BOOM economic scenario in Figure 3. Comparing a Single Subject's Priorities as Certainty Estimates and Consistency Ratio [CR] for Response Latency and Self-Report Methods.

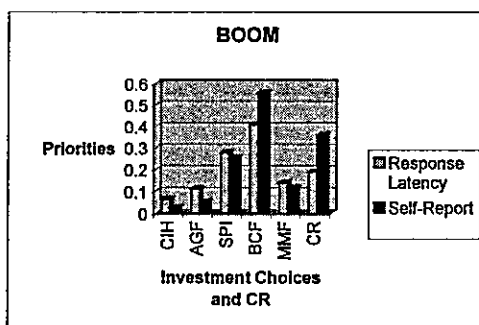


Figure 2. Comparing a Single Subject's Priorities as Certainty Estimates and Consistency Ratio [CR] for Response Latency and Self-Report Methods

A paired sample t -test is used to test the difference in the in the Certainty Estimates generated by each method. Since the research question focuses on the total difference, not the differences for one or more Certainty Estimates, a new variable is generated for each subject equal to the sum of the absolute differences between corresponding Certainty Estimates, for each method.

Thus, the total difference with respect to Certainty Estimates between the two methods for a single subject can be expressed as: $|CE_{11}-CE_{12}| + |CE_{21}-CE_{22}| + |CE_{31}-CE_{32}| + |CE_{41}-CE_{42}| + |CE_{51}-CE_{52}| = \text{CEDIFF}$. Where CE_{ij} = Certainty Estimate of choice "i" for method "j". There are a number of other distance measures that might be more appropriate for comparing the difference between two ratio scales [Saaty, 1999; Forman, 1999]; however, this measure was used for heuristic reasons. In another study, we plan to use other distance measures with the data collected for this effort.

A paired sample *t*-test is used employing CEDIFF, measured for each subject, as the new variable. Thus, the hypothesis that the methods produce different Certainty Estimates may be tested using CEDIFF. In a similar manner, a new variable called CDIFF is created that is equal to the difference in the Consistency Ratios for each subject.

Thus, the hypotheses are tested using paired sample methodology on the two new variables, CEDIFF and CDIFF. CEDIFF represents the difference in the Certainty Estimates for the response latency and self-report methods for each subject, while CDIFF represents the difference in the Consistency Ratios for the response latency and self-report methods for each subject.

RESULTS

Statistical Significance. Using paired sample *t*-tests, sufficient evidence was found to reject the null hypotheses ($\alpha = .025$) that the mean of the difference in the priorities between the two methods was zero ($p < .0001$). In a second test, there was sufficient evidence to be at least 95% confident of accepting an alternative hypothesis that response latency detects a lower consistency ratio than the self-report method ($p < .0001$).

Thus, it was found that the two methods produce different sets of Certainty Estimates and that the response latency method has a lower Consistency Ratio than does the self-report method. That is, when consistency is present, the response latency approach appears to detect it better than the self-report approach.

Practical Significance. In surveying the results, it seems as if a large number of response latency Consistency Ratios (CR_{RL}) are lower than the AHP .1 threshold for CR, and very few self-report CR_{SR} are beneath that threshold. This is a question of practical significance.

In examining the histogram of the Consistency Ratio for response latency, one finds a mean value of .094 and median of .087, both values less than the .1 threshold. More importantly, approximately 75% of Consistency Ratio samples for response latency fall between .119 and 0. However, the Consistency Ratio for self-report samples are distributed differently. The histogram for these samples has a mean of .315 and a median of .294, both significantly greater than the .1 threshold value. Further, only 10% of these samples fall between .123 and 0. For the self-report method, approximately 90% of the cases fall above the .1 threshold; however, for the response latency approach only 25% fall above the same threshold.

Thus, the differences found are not only statistically significant, there is some tentative evidence that the difference could be highly practical. Additional testing on a broader range of subjects and in different knowledge domains needs to be conducted in order to make any substantial claim in this area.

The AHP was selected as a suitable methodology to test the two approaches. However, the findings in this study offer evidence, tentative at best, that the response latency approach might hold promise as an additional technique that could be used with the AHP as another means of collecting decision data. Much additional work needs to be done such as testing the method using different measures of distance between the priority vectors, assessing the results using different time scales and thresholds, expanding the sample population to include experts as well as non-experts, testing the response-latency method against methods other than the one-to-nine scale for collecting decision data, and testing the assumption that a better approach is the one more sensitive to detecting consistency, when it is present, than the other.

REFERENCES

- Banagi, M.R., and Greenwald, A. G. [1994]. Implicit stereotyping and prejudice. In M. Zanna & J. Olson [Eds.], The Psychology of Prejudice: The Ontario Symposium, 7, 55-76. Hillsdale, NJ: Erlbaum.
- _____. [1995]. Implicit gender stereotyping in judgments of fame. Journal of Personality and Social Psychology, 68, 181-198.
- Cartwright, D. [1941] Decision-time in relation to differentiation of the phenomenal field. Psychological Review, 48, 425-422.
- Clemen, R. T. [1996] Making Hard Decisions; Duxbury Press, 265-298
- Crosby, F., Bromley, S., & Saxe, L. [1980]. Recent unobtrusive studies of black and white discrimination and prejudice: a literature review. Psychological Bulletin, 87, 546-563.
- Dashiell, J. F. [1937]. Affective value-distances as a determinant of esthetic judgment-times. American Journal of Psychology, 50, 57-67
- Dovidio, J.F., and Fazio, R. H. [1992]. New technologies for the direct and indirect assessment of attitudes. In J. Tanur [Ed.], Questions about Questions: Inquiries into Cognitive Bases of Surveys 204-237. New York: Russell Sage Foundation.
- Festinger, L. [1943] Studies in decision. Journal of Experimental Psychology, Vol. 32, No. 4, 411-423.
- Forman, E. [1999] Personal communications with Professor Ernest Forman, The George Washington University, School of Business and Public Management, Washington, DC.
- Gaertner, S. L., & Dovidio, J. F. [1986]. The aversive form of racism. In J. F. Dovidio & S. L. Gaertner [Eds.], Prejudice, Discrimination, and Racism 61-89. Orlando, FL: Academic Press.
- Harmon, P. and King, D. [1985] Expert Systems - Artificial Intelligence in Business; John Wiley & Sons, Inc. 41-43.
- LaBarbera, P. A., & MacLachlan, J. M. [1979]. Response latency in telephone interviews. Journal of Advertising Research, 19, 49-56.
- Marshall, K. T. and Oliver, M. R. [1995] Decision Making and Forecasting, McGraw-Hill, Inc. 240-241.
- Medsker, L. [1998]. Personal communications with Professor Larry Medsker, American University, Department of Computer Science and Information Systems.
- Nisbett, R. E., and Wilson, T. D. [1977]. Telling more than we know: verbal reports on mental processes. Psychological Review, 84, 231-259.
- Saaty, T.L. [1977]. A scaling method for priorities in hierarchical structure. Journal of Mathematical Psychology 15 79-84
- _____. [1999] Personal communications with Professor Thomas L. Saaty, the University of Pittsburgh.
- Sigal, H., and Page, R. [1971]. Current stereotypes: a little fading, a little faking. Journal of Personality and Social Psychology, 19, 247-255.
- Tyebjee, T. T., [1979]. Response latency: a new measure for scaling brand preference. Journal of Marketing Research, 16, 1, 96-101.