

THEORETICAL AND PRACTICAL ASPECTS OF AHP USING A SCALE DERIVED FROM THE TIME IT TAKES TO DECIDE BETWEEN TWO CHOICES INSTEAD OF ONE DERIVED FROM “1-9” ESTIMATES

Jerald L. Feinstein and Ross Lumley

School of Business and Public Management, Management Science Department
The George Washington University; Washington, DC 20052 – USA
jerry@gwu.edu, rlumley@gwu.edu

Keywords: AHP, decision time, response latency, market research, decision theory

Summary: *This paper discusses using response latency, or the time it takes for a subject to decide between two alternatives, to estimate relative levels of preference in the AHP. The response latency method is proposed as an alternative, in some cases, to the classic 1-9 scale used in the AHP.*

INTRODUCTION

This paper builds on the work presented at the 5th International Symposium on the Analytic Hierarchy Process, 1999 Kobe (Feinstein, J.L.; Winslow, E.; Nagy, T; and Coyne, J., 1999), where the idea was tested that the time it takes for a subject to decide between two alternatives in a set of pairwise comparisons, called the response latency approach, could be employed, in place of the classic approach of using a 1-9 scale of relative importance, to assess relative preferences in the AHP methodology. The faster a choice is made, the further the alternatives are separated on a ratio scale. Thus, distance in preference space is inversely related to some function of the decision time, or response latency as it is sometimes called.

Employing the self-reported 1-9 scale, subjects must indicate not only their preferred choice; in addition, they must provide an estimate of their preference for their choice compared to the other alternative. 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 preference for one choice over another is something that people find more difficult, as they are being asked to quantify something that they may not normally quantify in practice. To further complicate the situation, conscious processes are involved when deciding on relative levels of preference for controversial topics, and these processes are known to be vulnerable to manipulation. Subjects sometimes respond to questions in a manner that they view as socially correct, 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).

As a practical matter, if only the choice selected and the time it took to decide between the two alternatives could be used in place of adding the additional process of asking for a 1-9 assessment of preferences, then the questioning process is substantially accelerated, an unobtrusive measure is used that could potentially increase the criterion validity of the process, and if a higher consistency resulted, then the decision time would be a superior measure of relative preference in a number of situations.

As reported in the 1999 paper in Kobe, it was found that the two methods produce different sets of priorities and that the response latency approach results in a lower Consistency Ratio than does the classic method. That is, when consistency is present, the response latency approach appears to detect it better than the classic 1-9 approach.

In examining 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 the classic 1-9 self-report approach 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.

This information is illustrated and summarized in Figures 1, 2, and 3.

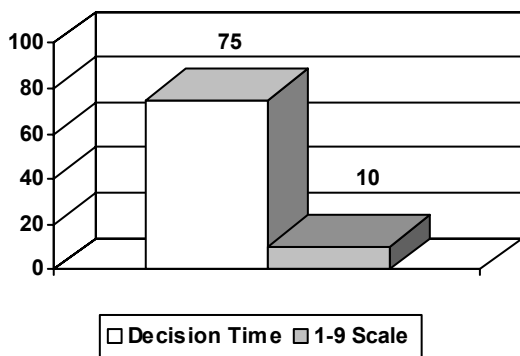


Figure 1 - 75% of the Decision Time Consistency Ratios Were Less Than .1 versus only 10% for the 1-9 Approach

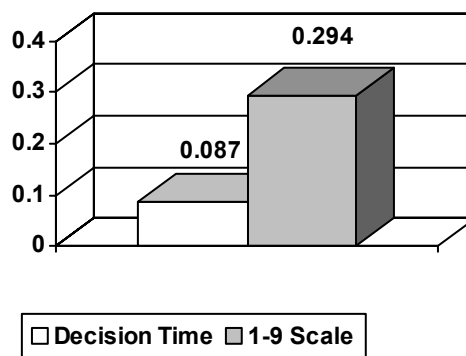


Figure 2 - The Median Consistency Ratio Was Substantially Lower for The Decision Time Method

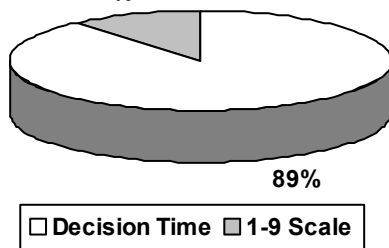


Figure 3 - The Decision Time Method Produced Lower Consistency Ratios (better consistency) 89% of the Time

However, the Kobe paper in 1999 and a journal article in 2000 (Feinstein, 2000) generated a number of questions that appeared to focus on two issues - - what might be the underlying model that supports the findings and a more detailed discussion of the more interesting anecdotal findings? This paper addresses those questions.

THE UNDERLYING MODEL

We believe the notion, that an inverse function of the decision time is a good estimator of relative preference, can be expressed in terms of sampling, classic hypothesis testing, and a number of assumptions about the underlying processes. For example, consider a subject who is faced with the task of choosing the heavier of two steel balls all of which have the same volume. The balls are picked up, one in each hand. If there is a very large difference in the weights, the choice is made almost immediately. As the balls get closer in weight, it takes longer for the subject to choose.

However, if one examines the subject closely, it is noticed that, for balls close in weight, each hand moves up and down in a very slight juggling motion. When one subject was asked what he was doing, the reply was, "I'm checking the weight." When the subject was "checking the weight," he appeared to be sampling the weight of the balls. The idea of subjective sampling is interesting as it can be taken further. Obviously data are being taken in the process, and the closer the weights of the two balls, the longer it takes to choose which one is heavier, and the assumption is that more samples are needed as the difference in the weights of the two balls approaches zero. Further, it is found that as the regret factor for making a wrong decision increases the decision time also increases in that more samples are needed to increase confidence. Thus, in general, during an eye exam, people take more time in assessing which lens combination is better than they might take in assessing which of two soft drinks tastes better. Thus, as increased confidence is required, the difference in the weight of the two balls decreases, or increased resolution is needed, the number of samples needed increases.

For each individual, a subjective sampling rate is assumed in terms of samples per second. At this point is not necessary to relate a specific number of samples per second to each individual. Rather, we assume that different people are "quicker" than others at making choices, and we correct for these time differences in our analysis. Measuring subjective sampling rates is a task for additional research.

In the previous case we discussed a type of subjective sampling that involved a subject checking relative weights of two steel balls using a slight juggling motion of the hands. In this case, one might argue that it is not too great a leap of faith to agree that some form of sampling is occurring within the subject's nervous system. Now, consider the next case where the subject is asked to do a pairwise comparison of which sport car to purchase or for which candidate to vote. Where there may be no physical sampling as there was in the previous example, some form of subjective mental sampling of memory is occurring. However, the result is the same; that is, it takes much longer to choose between similar candidates or sport cars than between ones that the subject views as very different.

However, in some instances, the model breaks down. It is found that in certain situations, a response can be rapid because it is associated with a deeply felt attitude and may be thought of as a conditioned reflex. Fazio (1995) claimed that what we call attitudes can be viewed as associations in memory between an attitude and an attitude object as well as one's evaluation of the object. The strength of this association determines the accessibility of that attitude from memory. This led Fazio to characterize the attitude-accessibility mental construct as a process where attitudes are "burned" into our mental constructs to

different degrees, based on their importance to us, for rapid access. Associations based on strongly held attitudes or beliefs are more quickly assessable by conscious processes than are weakly held ones. Thus, short response latency can be associated with more strongly held beliefs.

Of critical importance to market researchers is predicting the tendency for a person to act according to stated preferences. Not surprisingly, many subjects state intents and preferences on which they later act much differently. Thus their actions do not reflect what they say they will do. Fazio found that a good estimator of a subject's tendency to act in accordance with stated preferences is low response latency. When attitude strength is assessed using response latency, short latency equates to a strongly held attitude and long latency is associated with a weakly held one (Fazio, 95). Fazio claims that only strongly held attitudes tend to influence behavior, and that those attitudes are associated with very low response latencies. Fazio and Williams (1986) assessed preferences for presidential candidates.

A similar process occurs in database systems where critical information can be indexed for rapid retrieval. The phenomenon that Fazio is researching is important within the context of our research because it links short response times to high levels of preference. However, sampling and required confidence are no longer factors in the analysis. An example of this would be asking a conservative republican whether he prefers Clinton or Bush. However, in a choice between two conservative candidates the subject would revert to mental sampling.

Thus we have provided a brief description of what kinds of underlying processes might be involved that illustrate why the decision time works so well in scaling preference, with the added benefit of being able to predict the likelihood of action based on a subject's preferences. The proposed model is founded on a computer-based, information-processing paradigm. Elements of the model include sampling and subjective hypothesis where sensory input and stored memories are sampled, critical information elements are indexed for rapid retrieval, and various degrees of confidence are required to authenticate the final decision. In this context, as relative preference between two choices converge; confidence or certainty in the choice selected is reduced. As the required confidence thresholds increase - - that is as the importance of being correct increases, the decision time also increases. This explains why the decision time is greater for choosing between two unpleasant alternatives than choosing between two pleasant ones (Shipley, Norris, and Roberts; 1946)

SEVERAL INTERESTING EXAMPLES

The examples illustrated are drawn from research where the alleged goal was to identify the preferred mutual fund given two economic scenarios - - the first being complete global economic collapse, and the second being global economic expansion and prosperity. Within each scenario, subjects chose, in pairwise comparisons, the best alternative with the options being: (1) Cash In Hand (CIH), (2) Asian Growth Fund (AGF), (3) S&P Index Fund (SPI), (4) Blue Chip Fund (BCF), and (5) Money Market Fund (MMF). The choices were presented on a computer monitor, subjects selected the preferred choice, and the time or response latency was recorded. After each choice, the subject was asked to self-report, on the classic 1-9 scale (sometimes referred to as the verbal or self-report approach), the degree to which the alternative selected was preferred to the other. Thus, two measures were collected and compared - response latency (the decision time) and the verbal or self-report (classic 1-9 scale).

While the general findings are discussed in the introductory section of the paper, the following discussion and figures illustrate some of the more interesting findings. In all of the following figures the five investment choices (CIH, AGF, SPI, BCF, and MMF) and one consistency ratio (CR) are listed across the base of the bar chart. For the investment choices, the "Y" axis represents the magnitude of the calculated priorities and for the magnitude of the CR. The two conditions are shown side-by-side, with the response latency bar on the left and the verbal / self-report or classic 1-9 scale on the right.

Figure 4 illustrates an example of possible conscious censure from the global economic collapse or “Bust” scenario. It was proposed that the response latency technique was not susceptible to conscious censure for several reasons. The first is that subjects are unaware that their decision time is being recorded, and the second is that the underlying sampling mechanism may not be influenced by conscious censure. We propose additional research where in one set of trials subjects are told that their decision times are being recorded to assess their strength of relative preference, and in another set of trials they are not. If there are no practically significant differences in the results, then we might conclude there is some evidence to support the idea that the underlying mechanisms are not influenced by that knowledge.

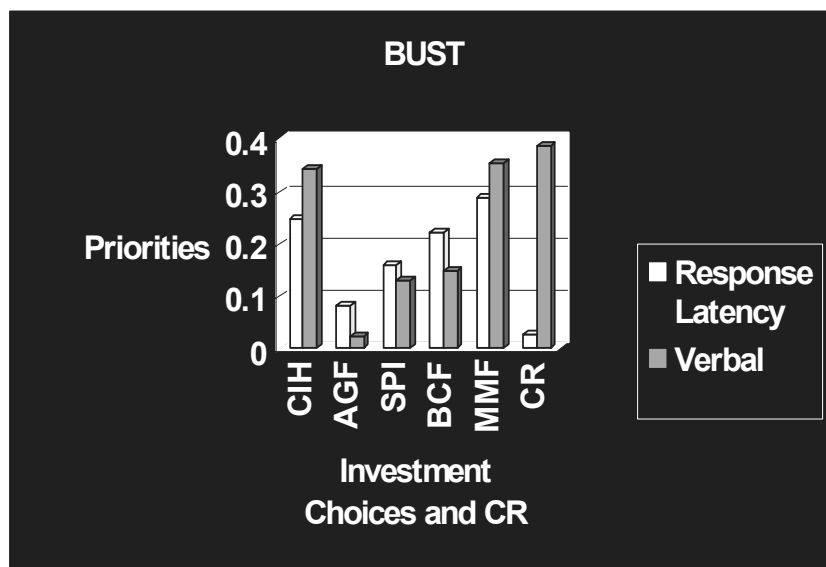


Figure 4 – Possible Case of Conscious Censure

The subject was interviewed after the experiment due to the unusual results obtained – the CR was very low for the response latency method, and very high for the classic 1-9 assessment. The subject claimed that he was a military man, and that once he made a choice, it was the “right” choice, so he gave a “9” as the relative preference level in all cases. However, when one analyzes the response latencies, one finds highly consistent results. Where similar results were obtained in other cases, this one was striking in its magnitude and clear reason for the differences in observations. In some other cases, subjects had different interpretations of the 1-9 scale even though they were briefed on the technique and claimed they understood it prior to the experiment. Others complained that they just do not think that way when making a decision.

There were a significant number of results where the two methods matched very closely, and both had acceptable CR values. Figure 5 illustrates the situation very well. It is interesting to speculate as to whether subjects who exhibit this result are different in any way from those who do not.

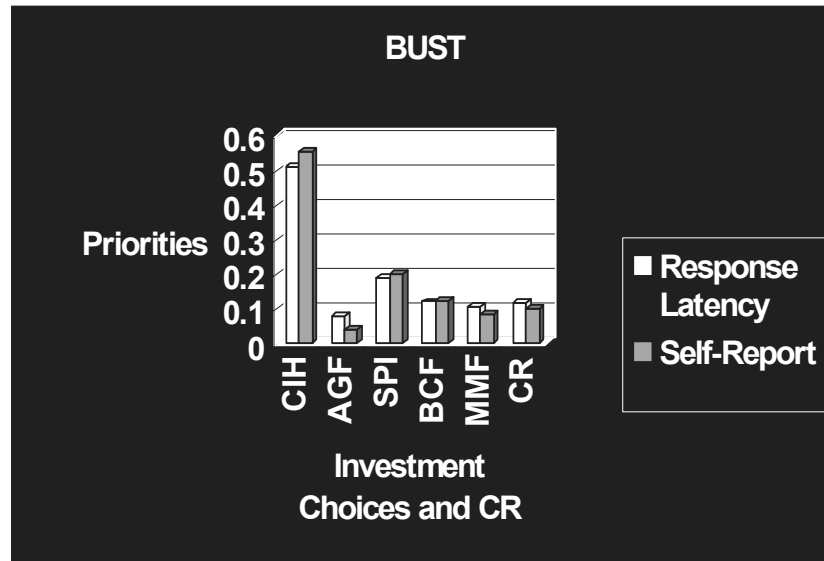


Figure 5 – A Subject Showing Almost Identical Results for Both

In the cases where the Self-Report/Verbal method produces a lower value for the CR, it is interesting to note that the value for the CR under the response latency approach is low also. Thus, we do not seem to obtain situations where low values of self-report/verbal CR are associated with high values for the response latency CRs. However, very often, low values for response latency CRs are associated with high values for self-report/verbal CRs. Figure 6 illustrates a case where a low CR value for the self-report/verbal approach is associated with an only slightly higher value for the response latency CR. In both cases, the values are at or less than .1

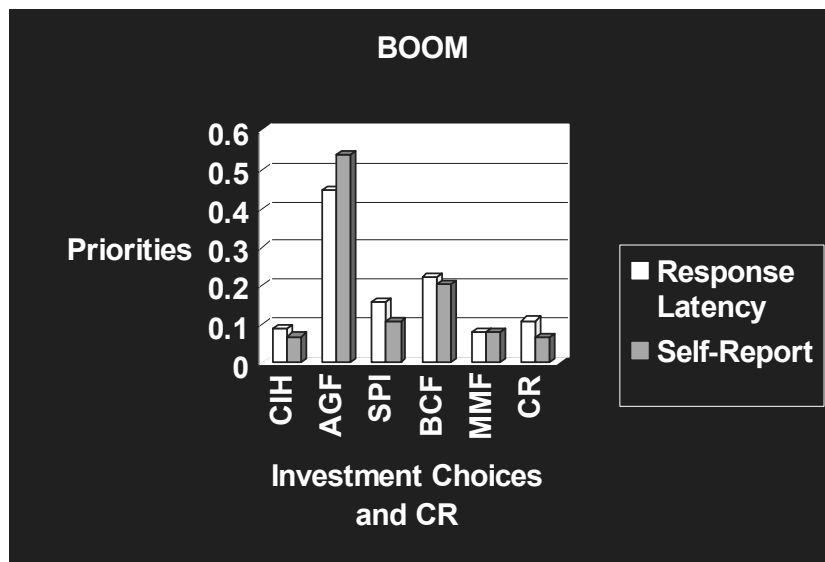


Figure 6 – Low Values for Self-Report / Verbal CRs Appear to be Associated with Low Values for Response-

In some cases, the response latency method produced extremely low CRs of which Figure 7 is an example. Again, it is interesting to speculate on the characteristics of subjects who produce such low CRs for the response latency case. In addition it is valuable to be able to harvest the additional consistency when it is present. Note in that in Figure 7 the CR for the classic approach was approximately .28 where it was near .01 for the response latency approach.

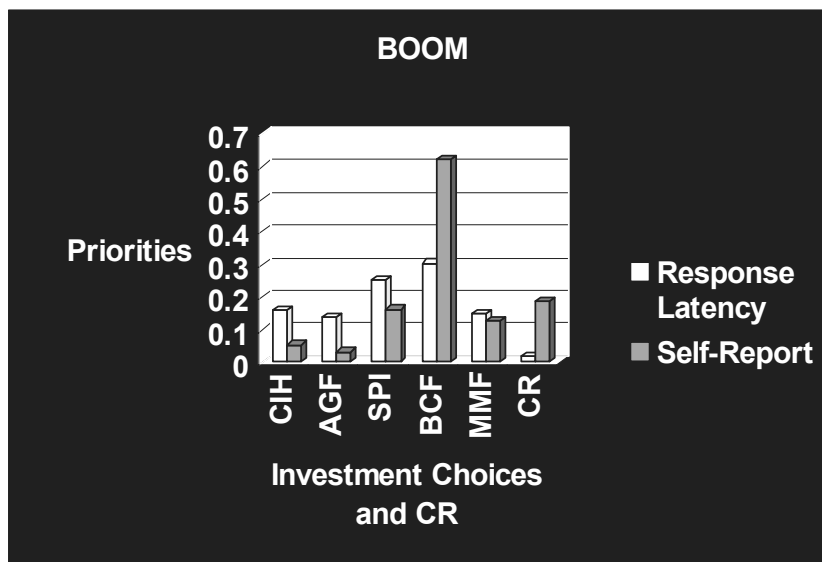


Figure 7 – An Example of a Very Low CR for Response Latency Compared to a Much Higher One for the Self-Report / Verbal Approach

CONCLUSIONS

We have shown that employing sampling models, hypothesis testing, and results derived from Fazio's research on attitude accessibility one can propose a reasonable model for the processes underlying the response latency approach. However, the model is only tentative, and requires additional research in the areas of subjective sampling and its relationship to the response latency approach. Further, in comparing the two approaches, we made the assumption that consistency among the preferences is a highly complex property and would not occur by chance; thus, the method used in comparing the two methods was based on another assumption that the approach which was a better detector of consistency, when present, was the superior one.

In analyzing some of the more interesting anecdotal findings, it was interesting to learn that some of the lowest CRs were derived from the response latency approach. In another example, we showed what we thought was a clear case of conscious censure appearing for the classic 1-9 approach and was absent for the response latency method. We wonder also why low CRs for the classic approach seem to be associated with reasonable ones for response latency; while the opposite is not true. Further, it is interesting to speculate on the nature of subjects who achieve very low CRs in either or both methods compared to those who score high. Finally, it has been proposed by others that since the response latency approach does not require the use of verbal language, particularly linguistic variables of preference, it could be used as a culture-free method to assess relative preference levels. These and other ideas will be explored in future research.

REFERENCES

- Banagi, M.R and Greenwald, A. G (1993). Implicit stereotyping and prejudice. In M. Zanna & J. Olson (Eds.), The Psychology of Prejudice: The Ontario Symposium, 7, 55-76. Hillsdale, NJ: Erlbaum.
- Greenwald, A. G, and Banagi, M.R., (1995). Implicit gender stereotyping in judgments of fame. Journal of Personality and Social Psychology, 68, 181-198.
- 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.
- 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.
- Fazio, R. H. (1995). Attitudes as object-evaluation associations: determinants, consequences, and correlates of attitude accessibility. 247-281. In R. E. Petty & J. A. Krosnick (Eds.), Attitude Strength: Antecedents and Consequences. Hillsdale, NJ: Erlbaum.
- _____ and Williams, C. (1986). Attitude accessibility as a moderator of the attitude-perception and attitude-behavior relations: an investigation of the 1984 presidential election. Journal of Personality and Social Psychology, 51, 3, 505-514.
- Feinstein, J.L., (2000) Comparing response latency and self-report methods for estimating levels of certainty in knowledge elicitation for rule-based expert systems, The International Journal of Knowledge Engineering and Neural Networks, Vol. 17 No. 5, 217 - 225
- Feinstein , J.L.; Winslow, E.; Nagy, T; and Coyne, J. (1999) Comparing response latency and self-report methods for estimating levels of certainty for rule-based expert systems employing the AHP Proceedings, 5th International Symposium on the Analytic Hierarchy Process Aug. 12 - 14, 1999 Kobe, Japan
- 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.
- 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.
- Shipley, W., E.D. Norris, and M.L. Roberts. (1946) The effect of changed polarity of set on decision time of affective judgements. Journal of Experimental Psychology, Vol. 36, No. 1, 237-243.
- Sigal, H., and Page, R. (1971). Current stereotypes: a little fading, a little faking. Journal of Personality and Social Psychology, 19, 247-255.