C. S. PEIRCE, CONFIDENCE AND RANDOM WALK THEORY

Stephen Link
University of California, San Diego. & McMaster University
link@ucsd.edu

Abstract

C. S. Peirce’s proposed measure of response confidence is shown to have a theoretical foundation in the random walk theory of discrimination. The paper reviews historical underpinnings of current research and uses new data to test the random walk predictions.

In yet another moment of inspiration, Charles Peirce (1877) observed that

… the chance of an event has an intimate connection with our degree of belief in it. …

There is a general law of sensibility called Fechner’s psychophysical law. It is that the intensity of any sensation is proportional to the logarithm of external force which produces it. It is entirely in harmony with this law that the feeling of belief should be the logarithm of the chance (pp. 708-709).

Investigating this proposal led Peirce, the father of American philosophical pragmatism, to enlist the aid of University of Pennsylvania graduate student Jacob Jastrow. In 1884 they reported their empirical findings in the Memoirs of the National Academy of Sciences. The experiment required a subject to discriminate changes in pressures applied to the index finger pad. Either the pressure increased from a base level and then decreased to it or, beginning at the incremented level the pressure decreased to the base level and then increased again. The subject was to determine which pattern of stimulation occurred. A portion of their data appears in Table 1. Confidence values (Marks) are summed across both stimulus patterns.

Peirce’s Law relates response probability and confidence through the Fechnerian idea that the strength of sensation is a logarithmic function of a ratio of stimulus magnitudes (Link, 1992). When applying Fechner’s idea to response probabilities, the numerator is a response proportion, p, and the denominator equals, (1 − p). That is, Peirce’s Law is expressed as,

\[ m = c[\ln(p/(1-p))] \]  

where \( m \) is the confidence, \( p \) is the response probability, and \( c \) is a scaling constant.

<table>
<thead>
<tr>
<th>Ratio of Weights</th>
<th>Response</th>
<th>Mark 1</th>
<th>Mark 2</th>
<th>Mark 3</th>
<th>Mark Total</th>
<th>Response Proportion</th>
<th>Average Confidence</th>
<th>Peirce's m</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.015</td>
<td>Right</td>
<td>110</td>
<td>51</td>
<td>3</td>
<td>1</td>
<td>165</td>
<td>0.660</td>
<td>0.324</td>
</tr>
<tr>
<td></td>
<td>Wrong</td>
<td>66</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>85</td>
<td>0.340</td>
<td></td>
</tr>
<tr>
<td>1.030</td>
<td>Right</td>
<td>106</td>
<td>72</td>
<td>23</td>
<td>2</td>
<td>203</td>
<td>0.812</td>
<td>0.548</td>
</tr>
<tr>
<td></td>
<td>Wrong</td>
<td>35</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>47</td>
<td>0.188</td>
<td></td>
</tr>
<tr>
<td>1.060</td>
<td>Right</td>
<td>86</td>
<td>75</td>
<td>54</td>
<td>24</td>
<td>239</td>
<td>0.956</td>
<td>1.040</td>
</tr>
<tr>
<td></td>
<td>Wrong</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Right</td>
<td>302</td>
<td>198</td>
<td>80</td>
<td>27</td>
<td>607</td>
<td>*estimated</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wrong</td>
<td>109</td>
<td>29</td>
<td>5</td>
<td>0</td>
<td>143</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ p = \frac{1}{2} \]

Peirce's \( m \) = 1.02, 1.92, 2.77, 3.97

\[ \ln[(n-.5)/.5] \]
This Law may be applied to results in Table 1 in several ways. First, a measure of direct confidence is the average confidence, based on confidence values ranging to 3 from 0, and shown in the last but one column of Table 1 for each of the three stimuli. Values of Peirce’s measure, m, appear in the last column. The comparison between these values is shown by the three dark circles in Figure 1. The two measures of confidence are linearly related. Peirce and Jastrow (1884) comment “The average marks seem to correspond to the formula.”

There is a second method of investigating these results that is more demanding. Each mark corresponds to a given direct measure for confidence values of 0, 1, 2, or 3. For each of these values the proportion of correct responses, and therefore, m is known. Applying the marginal correct response proportions for each Mark in Table 1 to equation (1) provides estimates of Peirce’s confidence to be compared with the direct values of 0, 1, 2, and 3. The amazing result, in Figure 1, is linear showing an excellent correspondence except for the non-zero intercept.

One could argue that the non-zero intercept is the result of a lack of continuity in the Marks themselves. The value zero may include cases where the confidence is as much as ½ or perhaps even more. Why should zero confidence correspond to anything other than guessing, which in this case gives a value of p equal to 0.50? What is your answer?

Subsequent research by V. A. C. Henmon (1911), Cattell’s student and Jastrow’s colleague at the University of Wisconsin, established relations between response confidence, response proportions and response time (RT). The study combines in a single experiment many desirable features, and stands as a monument to careful psychophysical experimentation. Three subjects were to discriminate between two horizontal line segments measuring 20 and 20.3 mm in length and were to rate their confidence in their judgment on a four part scale where A represented “perfectly confident,” B stood for “fairly confident,” C meant “little confidence,” and D “doubtful.”

Figure 1. Two tests of Peirce’s Law using Jastrow’s judgments in Table 1. Average confidence is the average of marks for each of three stimulus pressure ratios, 1.015, 1.030, and 1.060. The conditional confidence is based on the marginal proportion of correct responses given a confidence mark of 0, 1, 2, or 3.
Figure 2 shows subject BL’s mean correct and error RTs and correct response proportions for each confidence category. The proportion of correct responses increases steadily as the subject’s confidence increased. The proportion of times subject BL used each confidence category is shown by the black squares. Approximately one third of the responses are in the “Perfect” and “Fair” categories. Fewer responses are of “Little” confidence and very few “Doubtful.” With respect to response time, as confidence increases response time decreases. Error responses are faster than correct responses in each response category. A very surprising result is that the mean RT for errors is greater than the mean RT for correct responses.

Peirce’s measure computed for response categories D, C, B, A, yields values of -0.08, 0.63, 1.67, and 3.34. In this case the assignment of numbers to the confidence categories is somewhat arbitrary. Certainly no confidence is represented by a value of zero. But the unknown size of steps from “Doubtful” to “Little” to “Fair” to “Perfect” precludes the assignment of number to the other confidence categories. Still, Peirce’s measure increases dramatically from Doubtful to Perfect confidence.

A remarkable study by D. M. Johnson (1939) combined previous measures of performance with several innovations. First, Johnson’s subjects judged differences between horizontal line segments varying to 60 mm from 40 mm against a standard of 50 mm. But, the subjects received instructions to respond as Rapidly as possible without sacrificing accuracy, or to respond as Usual, or to make judgments as Accurately as possible without regard to speed. Following a choice response as to whether the comparison line segment was longer or shorter than the standard, subjects marked their confidence by a pencil mark on a 10 cm line segment.

**DATA FROM SUBJECT BL - HENMON (1911)**

![Graph showing mean response times and proportions across confidence categories.](image)

Figure 2. Mean response times shown on the left-hand ordinate. Proportions are scaled on the right-hand ordinate. Confidence categories are equally spaced on the abscissa. Results based on 1000 trials obtained at 50 trials per day.
Johnson’s experiment led to the Ph D thesis of David Ascher (1974). In “A model for confidence judgments in choice tasks” Ascher proposed that “the basis for confidence judgments in choice reaction time tasks is the distance between the initial and terminating values of the random walk decision process.” To test this idea Ascher employed the RT deadline method described by Link (1971). On each trial subjects viewed for 500 msec a fixed standard horizontal 20mm line segment displayed on computer controlled Tektronix X-Y display with black/white phosphor placed 1 meter away. Following a 200 msec delay a comparison horizontal line segment of 20, 19, 18, or 16 mm replaced the 20 mm standard. The 0Δs comparison occurred on 50 percent of randomized trials and 1, 2, and 4Δs on 1/6 of trials.

The subject’s response, “Same” or “Different,” and confidence were measured simultaneously. This measurement was accomplished by using a response panel with a Trial Initiation Key depressed by the subject to initiate a trial, and released to begin a response. Six response micro-switches arranged in an arc 3.5 inches distant from TIK were assigned confidence categories of High, Medium, Low, Low, Medium, High from left to right. For some subjects the three left-hand response buttons were used to indicate the response “Same” and the three right-hand buttons used to respond Different. Thus a subject could indicate High confidence in a “Same” judgment by depressing the left-most response micro-switch. Subjects were balanced across using the left or right-hand sides to chose “Same” or Different.”

Response speed deadlines of 350 or 475 msec appeared on the Tektronix X-Y display screen for 700 msec before the onset of the standard 20 mm line segment. Deadlines varied randomly from trial to trial. Two blocks of 310 trials were run per day. The first ten trials of each block were treated as practice and do not enter into the 600 trials per day analyses reported below.

The theory of speeded responding Link (1978a) proposed that RT deadlines caused subject’s to change response thresholds to meet speeded response demands. The more strict the RT deadline the nearer the response thresholds must be to the starting position value for accumulated information in order for responses to occur, on the average, within the deadline.

![Graph](image)

Figure 3. Mean response times averaged across RT deadlines of 350 and 475 msec for subjects JW and EL, 9600 observations total. White circles, mean RT “Same,” black circles mean RT “Different,” both scaled on the left ordinate. Numbers of observations scaled on the right-hand ordinate. Data from David Ascher Experiment 1 (1974).
Stricter RT deadlines caused subjects to place response threshold nearer the starting position resulting in faster responses and, according to Ascher’s hypothesis, the less confidence.

Average response times and numbers of responses for each stimulus difference, \(\Delta s\), appear in Figure 3. For 0\(\Delta s\) there are 4800 observations of which 3221 are correct, yielding a probability of a correct “Same” response of 0.671. The probabilities of correct “Different” responses increase from 0.546 for 1\(\Delta S\) to 0.720 for 2\(\Delta S\) to 0.861 for 4\(\Delta S\). White circles show the mean RT for “Same” judgments computed by dividing the sum of the mean RTs for Subjects JW and EL by two: these are 356 346 289 and 253msecs for 0\(\Delta s\) 1\(\Delta S\), 2\(\Delta S\) and 4\(\Delta S\). Mean RTs for “Different” judgments, black circles, vary from 300 to 353 to 361 to 344 msec.

In many respects these results extend the previous work of Henmon and Johnson to judgments of “Sameness” and “Difference.” The marginal mean RTs for 0\(\Delta s\), 1\(\Delta S\), 2\(\Delta S\) and 4\(\Delta S\) are 338, 349, 341, and 333msec for respectively. Note that mean error RTs are less than mean correct RTs and that this difference increases with stimulus difference. The average data provide a fair representation of the pattern of results within each RT deadline condition where mean RT for the 350 deadline equaled 299msec and for the 475 deadline 381msec. All trials enter into the analysis whether the RT deadline was met or not.

But our primary concern is with the confidence judgments. Subjects used three confidences for each choice (Same or Different), and therefore six possible categories of response. The split at the middle of the response panel into two choice categories divides the confidence judgments at the zero point. Confidences equally distant from the zero point can be considered to be of equal but opposite sign. Combining this idea with the supposition that the confidence categories are separated by equal intervals suggests a numerical assignment to confidence responses. In particular, values of \(-5, -3, -1\), were assigned to “Same” High, Medium, and Low confidence responses, and \(+1, +3, +5\) assigned to Low, Medium and High “Difference” confidence responses. Using these numerical assignments each subject’s average signed confidence judgments were computed for each stimulus difference and each RT deadline.

The Relative Judgment Theory of the psychometric function (Link, 1978b) shows that the probability of a response depends on the distance to the threshold for that response and a measure of stimulus difference, \(\theta\). These two parameters are estimated the using equation,

\[
\theta_A = \ln[p/(1-p)]
\]

where \(p\) is the probability of the response assigned to the response threshold.

Equation (2), the Logistic equation (Berkson, 1944), is a general consequence of bounded random walk theory (Link, 1978b). The obvious correspondence between equations (1) and (2) suggests that Peirce’s inspiration has a more formal foundation. The test comes from the results of Ascher’s experiment. Both stimulus difference, which affects \(\theta\), and the response thresholds, which are thought to affect the value of \(A\), must change from trial to trial. The value of \(A\) can be adjusted by the subject from trial to trial (Link, 1971; Link, 1992) to meet the speed requirement presented at the start of each trial, but the subject is initially unaware of the stimulus difference which should change the value of \(\theta\).

The independent measure of response confidence is the average signed confidence based on numerical values ranging to 5 from \(-5\). The two subjects, four stimulus differences and two RT deadlines provided 2x4x2=16 estimates of \(\theta_A\), Peirce’s measure, and measures of average confidence (one missing value due to a failure to obtain responses in all confidence categories). These values are plotted shown in Figure 4. The correlation between these values is 0.976. The close correspondence between these two measures confirms our expectations.
RESULTS BASED ON ASCHER

○ EL350  ● EL475  □ JW350  ■ JW475

$R^2 = 0.9524$

PEIRCE’S MEASURE

-4  -3  -2  -1  0  1  2  3  4

SIGNED CONFIDENCE

Figure 4. Estimates of θA, Peirce’s measure, and averaged signed confidence based on numerical assignments described in the text. Averages based on values ranging from -5 for High confidence “Same” responses to +5 for High confident “Different” responses.

In conclusion, Peirce’s confidence measure, m, now has a new and appealing theoretical interpretation. Both stimulus difference and the response threshold, A, which is a measure of the amount of information required for a response, jointly affect a subject’s confidence.

References


Henmon, V. (1911) The relation of the time of judgment to its accuracy. Psychological Review. 18, 186-201.


Fechner Day 2003

International Society for Psychophysics
Larnaca, Cyprus, 2003
Fechner Day 2003

Proceedings of the Nineteenth Annual Meeting of the International Society for Psychophysics
Larnaca Bay, Cyprus, October 19 – 23, 2003

Editors

Birgitta Berglund & Elisabet Borg
Department of Psychology, Stockholm University, Sweden

International Society for Psychophysics
Stockholm, Sweden, 2003
EXECUTIVE COMMITTEE:
Stephen Link (President)
University of California at San Diego, CA, USA

Galina Paramei (Treasurer)
University of Magdeburg, Germany

Somon Grondin
Université Laval, Canada

Wolfgang Ellermeier
Aalborg University, Denmark

Birgitta Berglund
Stockholm University & Karolinska Institutet, Sweden

HISTORIAN:
Helen Ross
University of Sterling, Scotland

LONGTERM ACCOUNT HOLDER:
Debra Zellner
Montclair State University, NJ, USA

ORGANIZING COMMITTEE:
The Gösta Ekman Laboratory
Stockholm University & Karolinska Institutet, Sweden

Birgitta Berglund
Elisabet Borg
Demetris Demetriou
Ingegerd Johansson
Britth Sandin
CONTENTS

Actis-Grosso, R. & Vezziani, S.
Rediscoveries in perception: toward a database ............................................ 1

Alberdi, I.
Choosing a vacation spot using Borg’s scale: A collective action perspective .......... 7

Algom, D., Eidelis, A. & Townsend, J. T.
Stochastic modeling of stroop stimuli processing: A puzzle .............................. 13

Alves, N. T., Fukusima, S.S. & Da Silva, J. A.
Perception of facial expressions of emotion and brain lateralization ...................... 17

Ben-David, B. M., Levy, L. & Algom, D.
The emotional stroop effect is a generic reaction to threat, not a selective reaction to specific semantic categories .......................................................... 21

Berglund, B., Johansson, I. & Zheng, L.
Odor and sensory irritation perception in women with chemical sensitivity ............ 25

Borg, E.
The multidimensional character of perceived exertion for a maximal and sub-maximal cycle ergometer test .......................................................... 31

Borg, G.
An index for relations between perceptual magnitudes based on level-anchored ratio scaling ..... 37

Casco, C., Grieco, A. & Giora, E.
The relation between saliency of motion and form .......................................... 43

Clara, E., Agrillo, C., Regolin, L. & Vallortigara, G.
Spontaneous preference for symmetrical stimuli in visually na"ive domestic chicks .......... 47

Cools, W., Hofmans, J., Baekelandt, S., Mairesse, O., Strubbe, M. & Theuns, P.
The role of context in Category Scaling: Placing anchors at equal distances .......... 53

Ehm, W. & Wackermann, J.
Analysis of a stochastic version of the ‘klepsydra model’ .................................. 59

Eisler, A., Eisler, H. & Yoshida, M.
Perception of global risks: A human ecological approach .................................. 65

Ellemeier, W., Narens, L. & Dielmann, B.
Perceptual Ratios, differences, and the underlying scale .................................... 71

Esteban Alberdi, C., Rubio Andrade, I. & Rodriguez Anton, J.M.
Client’s perception of the service received in travel agencies ................................ 77

Fitousi, D., Chajut, E. & Algom, D.
Automaticity versus attention in the processing of two-digits numerals ................ 83
Monahan, J. S. & Sheffer, S. M.
The vocal reverse stroop effect: Interference and congruity .............................................. 187

Moreno, F. J. & Peralta, H.
Familiarity, visual complexity and age of acquisition bias: The assessment of semantic
deterioration in Alzheimer’s patients .................................................................................. 193

Nilsson, M.E., Berglund, B. & Axelsson, Ö.
Soundscapes perceived by children and adults .................................................................. 199

Nilsson, M.E. & Nougo, A.
Noise sensitivity and soundscape perception ...................................................................... 205

Oberfeld-Twistel, D.
Modeling the effects of forward-masker level on the loudness of a tone ............................ 211

Do valence and arousal prompt the same or different ERP components? Pointing towards an
inner psychophysics ........................................................................................................... 217

Paramei, G., Hamada, J. & Ehrenstein, W. H.
Size induction: A Judgment-order effect .......................................................................... 223

Pavlova, M., Sokolov, A., Staudt, M., Birbaumer, N., Enck, P. & Krägeloh-Mann, I.
Engagement of periventricular parietal regions in biological motion processing .................. 229

Petrusic, W. M.
Calibration of response times and confidence in perception and cognition .......................... 235

Plank, T. & Ellermeier, W.
Discrimination of temporal loudness profiles .................................................................... 241

Regolin, L. & Vallortigara, G.
Aspects of visual perception and cognition in the avian brain ........................................... 245

Rondonini, C. & Galera, C.
Stereooscopic sensitivity and stimuli selection in visual search .......................................... 251

Ross, H. E.
Context effects in the scaling and discrimination of size ..................................................... 257

Rusiak, P., Lachmann, T. & Jaskowski, P.
Mental rotation of letters in dyslexics and implications for diagnosis and educational
treatment ........................................................................................................................... 263

Sarris, V.
Auditory psychophysics in the light of perceptual-cognitive context effects ......................... 269

Sartori, R. & Tommasi, M.
The effect of prior stimulus information upon category ratings ............................................ 275

Scharf, B., Nieder, B., Buus, S. & Cazals, Y.
Induced loudness reduction: Correlations between ears and procedures ............................ 281
Servos, P.
Mapping the cortical substrates of biological motion perception: Associations and dissociations ................................................................. 287

Shaki, S. & Petrusic, W. M.
Instruction interference and the semantic congruity effect ................................................................. 293

Shendyapin, V. M. & Skotnikova, I. G.
Formal neurons approach for sensory judgments modeling ................................................................. 299

Sokolov, A., Pavlova, M., Lutzenberger, W. & Birbaumer, N.
Multisensory processing and meg gamma-activity .................................................................................. 303

Teghtsoonian, M. & Teghtsoonian, R.
Putting context effects into context .................................................................................................. 309

Townsend, J. T., Fific, M. & Assadi, A.
General recognition theory and probabilistic perceptual separability on simple cognitive surfaces ........................................................................... 315

Travieso, D. & Suero, M.
Psychophysical functions of the combining effects of Müller-Lyer and horizontal-vertical illusions ........................................................................... 319

Vickers, D.
Confidence and response time in three-category judgment ................................................................. 325

Wackermann, J., Ehm, W. & Spliś, J.
The 'klepsydra model' of internal time representation .......................................................................... 331

Ward, L. M.
Metapyschophysics: Some “Big” questions and tentative answers .......................................................... 337

Zimmer, K. & Baumann, O.
Validity of Luce’s transformation function relating numerals and numbers in loudness fractionation ............................................................................. 343