Chapter 5

Results of Experimental Study

Overview: Unfortunately, only 79 subjects produced usable and identifiable data in both parts of the experiment (the pre-test and the computer-based exercise). Given 138 participants, we expected to have seven or eight more subjects per cell in our analysis (after reasonable losses). However, we remain skeptical that these additional subjects would drive our non-significant relations to significance. First of all (and speaking generally), we did see some evidence of schema-driven processing in our subjects. For instance, the most peculiar or counter-intuitive feature of the H-SAN model may be the hypothesized lack of direct inter-item connections between congruencies. This has been supported previously by demonstrating longer inter-recall intervals between successive congruent recalls (e.g., Srull, 1985). As you can see from the graph below (Figure 11), this is likewise evident in our data (although the time intervals between incongruencies and irrelevancies were strangely brief). We should caution: many of the longest C-C intervals came late in some subjects' recall performances. These lengthy time intervals may be due to time wasted in re-visiting prior recalls. Other evidence of H-SAN processes are mentioned in more detail below as they pertain to specific hypotheses. At this point, the reader should just note that there is some evidence that (counter to the H-SAN model) neutral-irrelevancies are clustered together. Their inter-recall intervals are brief, and they are recalled in sequential (R to R) pairs at a rate almost double that of similar pairs involving irrelevancies (e.g., I to R).

Before testing specific hypotheses, we continued to look for evidence of person-memory schema-based processing in general (per the ANCOVA and ATI analysis described in the preceding chapter). Since our hypotheses focused on the early blocks of presented sentences (as this is where incoming prejudices are likeliest to make a difference), we first calculated relative probabilities for recollections of such sentences.

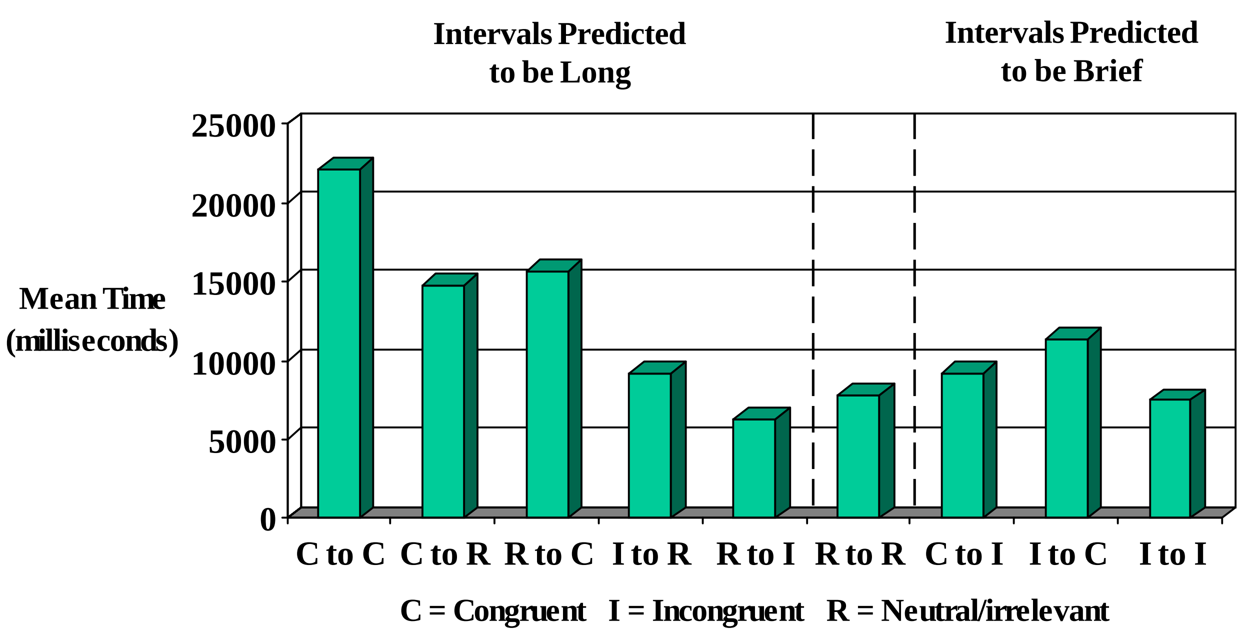

Figure 11. Average Inter-recall Response Intervals by Sentence Type

Given three types of sentences (I=inconsistent, C=consistent, and iR=neutral/irrelevant), there were nine possible inter-recall intervals. Our subjects averaged 9.69 recollections from the first four blocks (i.e., the first half of the stimulus material). So an average subject would produce nine intervals — and given nine types — the probability of "drawing" the first of one type would approximate the overall probability for that type.

Since our manipulation included disparate numbers of sentence types (I=12, C=8, and iR=8 in the first four blocks), we would expect observed inter-recall intervals – in the absence of schema effects — to appear in numbers predicted by simple probabilities adjusted for these disparate numbers (i.e., primacy and recency effects should balance out given random sentence presentation order across subjects). Since the subjects are told that duplicate (i.e., redundant) recollections do not count, the random-process probabilities calculated assumed sampling without replacement. For example, based on the relative proportions of sentence types just given, the probability of recalling a C followed by another C would be {12/28} times {11/27} equals 0.1763 (absent schema effects). In other words, P(CC) = 0.1763, P(iRiR) = 0.074, P(CiR) = 0.127, P(iRC) = 0.126, P(II) = 0.074, P(IC) = 0.126, P(CI) = 0.127, P(IiR) = 0.0844, and finally, P(iRI) = 0.0844 (again, all of these assume the absence of schema effects and the minimization – via random sentence sequences — of primacy and recency effects).

Summing these for the combinations favored by schema effects, we get a combined probability of P(II or IC or CI) = 0.327; conversely, for those not favored, P(CC or iRiR or CiR or iRC or IiR or iRI) = 0.672. Of course, these two probabilities sum roughly to unity – and ideally, we would hope to find that our subjects' recall exceeded the former and failed to reach the latter (at least for those subjects in the Black target condition). Such an outcome would support the notion that schema effects are present – a premise of our more specific hypotheses. While technically these effects were manifested in the predicted directions (in the Black target condition), they were trivial in size (i.e., quite insignificant). Specifically, observed P(II or IC or CI) equaled 0.336 and the observed P(CC or iRiR or CiR or iRC or IiR or iRI) equaled 0.664; these are both very close to the above numbers we'd predicted for random processes (our first hint that inconsistency resolution might not be prevalent). Fortunately, when using the White target data alone, these random probabilities were nearly nailed to three decimal places (and from the other direction). The good news here was that more inconsistency resolution did seem to be going on in the *Black target fast presentation condition* (observed P(II or IC or CI) equaled 0.357) and least of all in the *Black target slow condition* (observed P(II or IC or CI) equaled 0.309) where bolstering hopefully took its place. However, these differences were not statistically significant ($t=.949$, $p = .347$) . In general, recall proportions were weak indicators in our study (largely due to vast amounts of across subjects variability unaccounted for in our model). The inter-recall time interval data was somewhat more enlightening – possibly due to the subjects' presumed inability to knowingly corrupt it (though we can't imagine a corrupting filter that leaves the conditional probabilities random).

Our praise for the time interval data must not be overemphasized since our pre-test attribute measures (with the exception of one DV in one cell) predicted neither recall proportions nor time intervals at the individual level. This was an ambitious goal in the first place – akin to using a batting average to predict a batter's performance during a single (and lifetime unique) period of 91 to 126 seconds at bat. In other words, our Srull expectancy emergence charts (see Figures 19 to 22, pp. 84-87) suggest that general impressions were formed – even by low prejudice subjects – in the first 91 to 126 seconds of stimulus presentation. If prejudicial expectation (PE) scores were to predict differences in recall behavior, their only real shot at doing so was for recollections based on the first 91 to 126 seconds of stimulus sentence presentation. Apparently by that point, nearly all of our subjects had decided our fictional target was not an admirable student. Thus the distinctions (between high and low PE) that *were relevant to our study* became insignificant beyond this initial stimulus exposure time.

To add more confusion, our candidate covariate – the potential contaminate need for cognition (nCog) – failed to have consistent effects. In six significance tests (against *both* omnibus test individual level DVs in each of three cells of subjects), nCog was only significant in one of six (see Table 5 below). This was when predicting schema quickening estimate (SQE) in the Black target/9 second (fast presentation) condition. Elimination of outliers (guided by SPSS casewise analysis) failed to improve these outcomes.

Unfortunately, the nCog regression slopes also varied significantly from cell to cell. Thus, there is no straight-forward way to remove nCog's influence (as a latent construct) from our subjects' performances. It follows that linear relations between our continuous variables are largely beyond comprehension (i.e., the principal assumption of ANCOVA was not even marginally satisfied). Prejudice (PE) never achieved statistical significance in predicting schema quickening estimate. With congruency recall advantage (CRA) as the DV, PE only achieved statistical significance in the Black target/9 second (fast presentation) condition — and nCog failed to reach significance in any. So in testing our specific hypotheses (see below), we were forced to drop nCog from our model and dichotomize our continuous predictor PE.

Table 5. Regression outcomes for analysis of candidate covariate nCog.

Cell:	DV	B	Std Err	Beta	t	Significance
Minority target + 9 secs	CRA	-6.4E-03	.008	-.153	-.774	.446
Minority target + 13 secs	CRA	+8.3E-03	.013	+.145	.657	.519
Non-minority + 13 secs	CRA	+4.7E-03	.007	+.177	.649	.528
Minority target + 9 secs	SQE	-2.2E-02	.006	-.567	-3.442**	.002
Minority target + 13 secs	SQE	-9.8E-04	.009	-.026	-.115	.910
Non-minority + 13 secs	SQE	+1.5E-02	.017	+.229	.849	.411

In our own defense, we can say that the simple bi-variate correlations observed in our data set may have justified our obsession with controlling for ability differences. This obsession led to the complicated omnibus DV's: CRA and SQE. For instance, it would have been cleaner to test PE's ability to predict the probability of recalling congruencies (i.e., one might intuitively expect the more prejudiced subjects to better recall the more pejorative sentences in the Black target condition).

Table 6. Regression outcomes for analysis of PE predictions of CRA and SQE .

Cell:	DV	B	Std Err	Beta	t	Significance
Minority target + 9 secs	CRA	+1.8E-02	.006	+.514	2.997**	.006
Minority target + 13 secs	CRA	+4.5E-04	.011	+.009	.042	.967
Non-minority + 13 secs	CRA	- 1.6E-03	.017	-.028	-.098	.924
Minority target + 9 secs	SQE	+3.3E-03	.007	+.100	.505	.618
Minority target + 13 secs	SQE	-7.5E-05	.007	-.002	-.010	.992
Non-minority + 13 secs	SQE	-6.1E-03	.023	-.076	-.264	.796

The problem, of course, was that prejudice might be inversely related to verbal intelligence which was likely to be positively related to recall abilities of all sorts — including recall of pejorative congruencies. Our data were passably consistent with these concerns: the moderately negative correlation between PE and the summed verbal intelligence items from the Wonderlic was significant (r = -.25, p < .05, N = 79); likewise, the moderately positive correlation between the latter and total recalls was also significant (r = .21, p < .05, N = 83). Thus we created the ratios CRA and SQE so as to control for differences in recall ability and verbal processing speed. These omnibus DV's led to the disappointing results described above. We still felt it prudent to test our specific hypotheses — to which we'll turn after a brief discussion of assumption checks.

Assumption Checks: We used scatterplots and normal P-P plots of regression standardized residuals to check our assumption of linear relations between our predictors (PE and nCog) and our DV's (CRA and SQE). No obvious relations of any kind emerge from viewing these plots. Normal Q-Q plots of weighted average inter-recall intervals were also analyzed prior to testing our third hypothesis. These were produced for each cell and each category of interval. Moderate deviations from the normality assumption provoked us to repeat our test with appropriate non-parametrics. These yielded the same results.

Results Pertaining to Hypothesis One: This hypothesis focuses on "subjects with no prior expectancies concerning the target" that are assigned to the one cell where we applied "the *fast-stimulus presentation* condition" (p. 39, Chapter 4). Specifically, it suggests we should see a Srull expectancy-emergence pattern similar to Figure 6 (in the previous chapter) when sorting recalls from our low PE scorers (since low PE scorers should be less likely to hold an initial impression based on the target's race alone). This hypothesis is only partially supported (and even at that, what appears visually supportive is not statistically significant). In Figure 12, it is apparent that lacking an initial impression, subjects seem to process the three types of sentences in a similar fashion. Once an impression is formed, the sentences are processed in different ways apparently dependent on their type (I, C, or iR). We further hypothesized that under this fast condition, our subjects' recall would ultimately favor incongruencies. We'll deal with this issue again when we discuss computer monitor stimulus presentation (versus our predecessors traditional three-ring binder method of presentation), but for now.....

Looking at our actual results (see Figure 12 below), it is not clear this latter aspect of Hypothesis One finds support. However, our low PE (i.e., lowest third) subjects do appear to be treating the sentence types similarly at the outset. An expectancy seems to emerge in the third block of sentences (about 126 seconds into the sentence list). When Srull expectancy-emergence charts are presented for the other groups relevant to other hypotheses, this finding often repeats. In other words, even with as few as 14 sentences, most subjects have formed an impression in as little as 91 to 126 seconds. A variance comparison test for correlated samples (Guilford, 1965) was used to verify the visually-apparent increase in recall probability dispersion before and after expectancy emergence (i.e, impression formation). Variance for the first three blocks ($s^2 = 0.00476$) was compared to that for the latter three blocks ($s^2 = 0.00640$) given an early block to late block recall correlation of 0.473 ($p < .01$). Difference in variances was not statistically significant ($t = 1.0829$ vs. a critical $t = 2.306$).

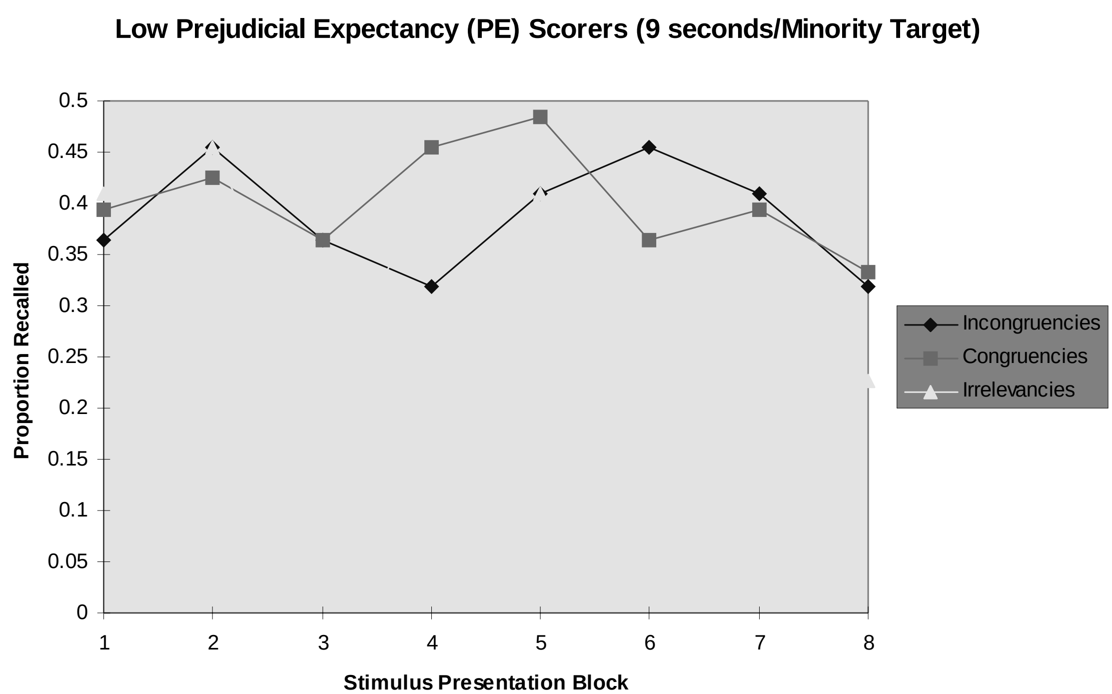

Figure 12. Srull Expectancy-Emergence Chart for Hypothesis One

Results Pertaining to Hypothesis Two: This hypothesis was focused on highly prejudiced subjects (high third of PE scores). Here, we expected to see results similar to Figure 7 (in the preceding chapter). Again, we have only partial support (and again, evidence that is visually supportive is not statistically significant). Using the high PE subjects from the same cell as used in the figure above, the contrast appears profound. As predicted, our high PE subjects appear to be processing the sentence types differently from the very start. This suggests support for our expectation that highly prejudiced subjects would form an initial impression of the target based largely on the target's race.

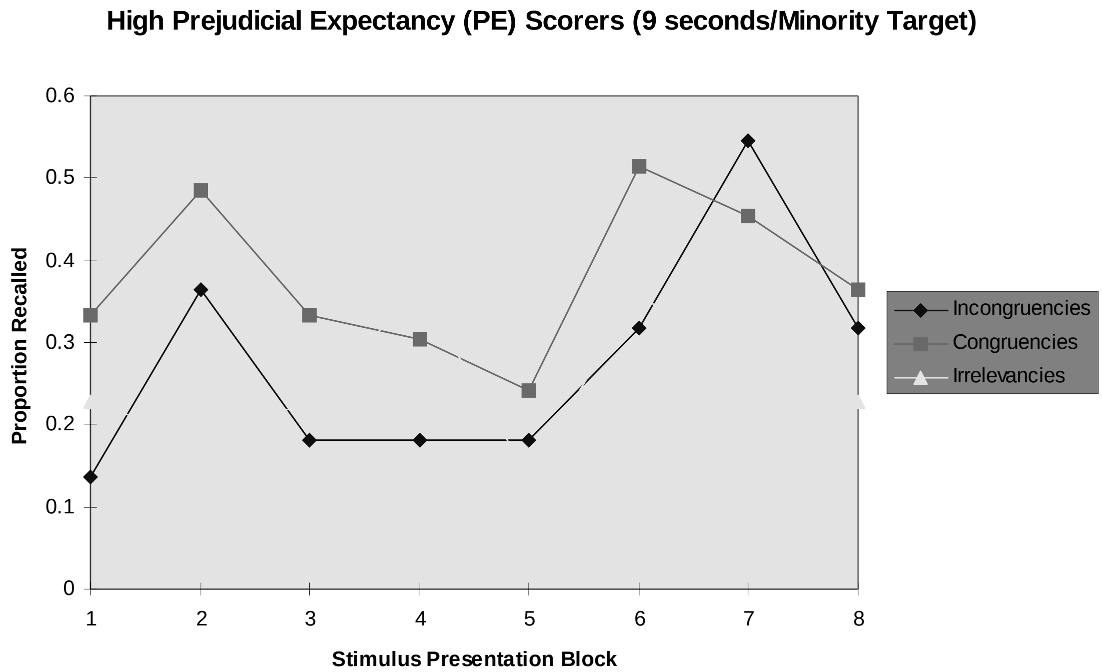

Figure 13. Srull Expectancy-Emergence Chart for Hypothesis Two

Before applauding too heartily, two points should be considered. First, we again predicted that ultimately these subjects would show enhanced recall for incongruencies (via increased emphasis on inconsistency resolution). It is not obvious that such occurred. To repeat, this may relate to computer monitor versus three-ring binder stimulus presentation (please see next chapter). Secondly, similar comparisons in the slow presentation condition are not so stark (see Figures 14 and 15 below). There are similarities, but even low PE scorers seem to form an impression in the first 91 to 126 seconds. Again, however, our highly prejudiced subjects treated the sentence types differently from the start. The problem is that our low prejudice subjects start out almost as bad — and catch up fast. A variance comparison test for independent samples (per Guilford, 1965) was performed in an attempt to confirm the visually-evident differences between low and high PE subjects. The variance of recall proportions for the first three blocks for high PE subjects ($s^2 = 0.0120$) was compared to that for our low PE subjects ($s^2 = 0.00476$). Again, the visually apparent difference was not statistically significant ($F = 2.521$ vs. a critical $F = 3.44$).

Results Pertaining to Hypothesis Three: This hypothesis was tested via paired sample t-tests within each cell based on the dichotomized form of our predictor PE. Though we had originally planned on direct statistical tests of nCog residuals across cells, this became problematic when nCog failed to perform consistently from cell to cell. To maintain statistical power, we used a median-split within each cell — allowing six paired comparisons. These compare the average *supposedly* schema-speeded inter-recall time intervals to the average of *supposedly* non-speeded intervals. This is repeated six times — within the high PE scorers — then again within low PE scorers — in each of our three cells.

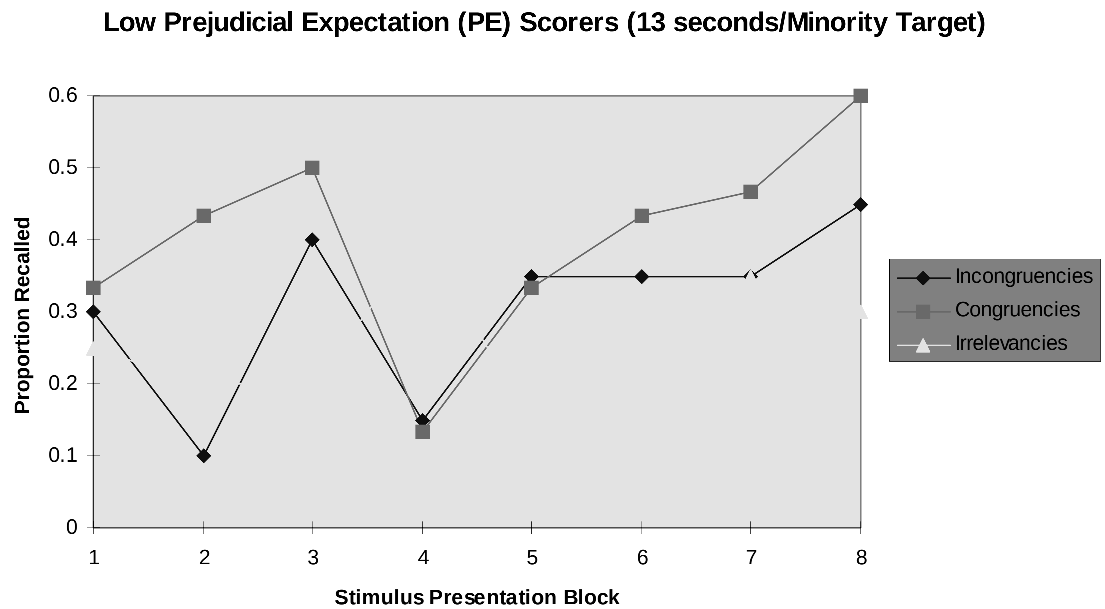

Figure 14. Srull Expectancy Emergence Chart

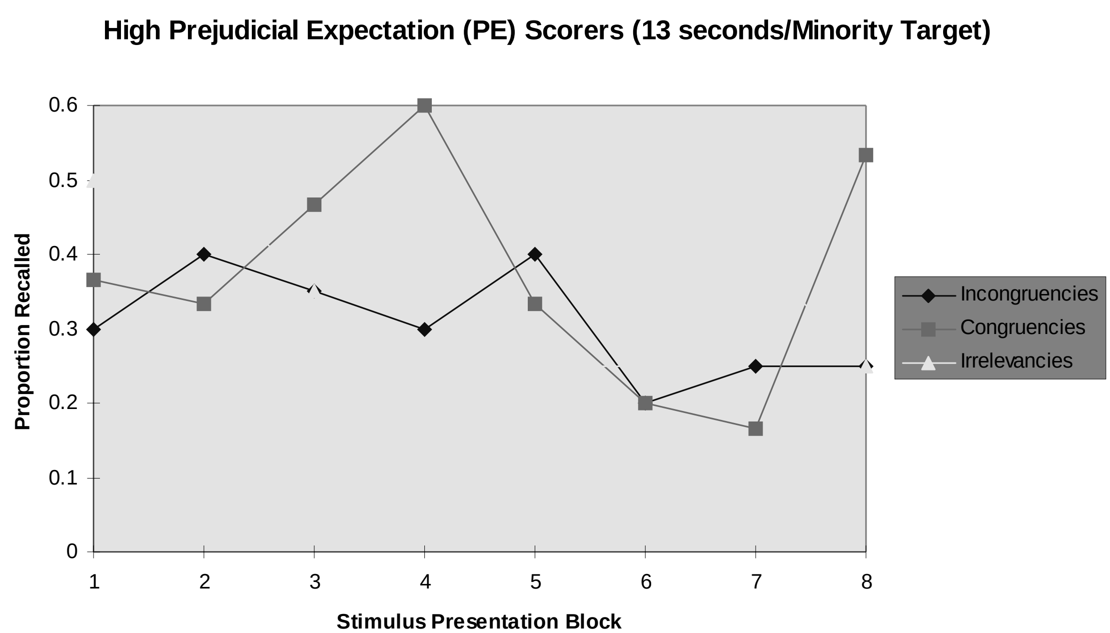

Figure 15. Srull Expectancy Emergence Chart

Remember that Hypothesis Three predicts "*average inter-recall response times* will be quicker between recollections of two inconsistent sentences (or a consistent and inconsistent sentence) than between two consistent sentences or between any sentences where at least one is irrelevant. Low racism scorers should not have formed an impression of the Black target in the first few blocks (i.e., the first 21 sentences), so evidence (in their data) of direct inter-item links between inconsistent sentences should be no greater than the equivalent data from similar subjects in the White target condition. However, inter-recall response times (associated with the early blocks of sentences) should indicate that high racism scorers *have* formed a negative impression of the Black target (possibly immediately), and this evidence should be absent for high racism scorers in the White target condition" (p. 40, Chapter 4).

As with the previous hypotheses, this one is only partially supported. Consistent with our above-mentioned finding for long intervals between successive recalls of congruencies, there is some evidence that inconsistency-resolutions or inter-item comparisons have spawned inter-item links — and that these links are more likely found between I's or between I's and C's (and not between any two C's). But, as with the other hypotheses, we cannot show support for inconsistency resolution holding sway against bolstering (or vice versa). This is because the manipulated presentation time (the supposed bolstering opportunity) does not appear to have made a statistically significant difference.

That said, the good news is that among the six paired sample t-tests described above, the only two that should have been significant indeed were the only two near significance. Specifically, using median splits (recall data associated with the first half of the stimulus set and subjects split into high and low PE scorers), the weighted average of schema-speeded inter-recall intervals was significantly faster than non-speeded intervals only for high PE scorers reading about a Black target. This is entirely logical if one assumes that H-SAN processes are relevant to racial prejudice. The interval differences do not always change substantively from cell to cell, but the H-SAN processes appear to constrain time interval variance where prejudice applies in ways consistent with person-memory schema. The specific mean differences (in milliseconds) and significance outcomes follow (see Figure 16).

If forced to draw conclusions from a single data set, the similarities seen in the White target cell suggest that a person-memory schema forms very quickly; however, in the presence of a well-practiced prejudicial expectation, it may form instantly. Thus, less variance is observed in the high prejudice subjects, making their time interval differences significant. In the relative absence of a prejudice, the schema may form at differing rates across subjects, yielding more variance and less of chance for significant time interval differences. Again (as with the aggregated recall probabilities mentioned at the beginning of this chapter), there is weak evidence suggesting that high PE subjects engage in more inconsistency resolution in the fast stimulus condition. In other words, the difference between speeded and non-speeded intervals is more than doubled comparing the high PE subjects between the 9 second and 13 second cells (10,698 vs. 5041 msec; however, this was also non-significant: $t = 1.35$, $p = .193$). This may weakly indicate the presence of more direct inter-item links resulting from a greater number of cross-comparisons (thought to be a manifestation of attempts to resolve inconsistencies).

	Black Target	White Target
13 seconds	High PE: $E(\mu_D) = 5041$ p = .002* Low PE: $E(\mu_D) = 5653$ p = .20	High PE: $E(\mu_D) = 6158$ p = .19 Low PE: $E(\mu_D) = 4462$ p = .14
9 seconds	High PE: $E(\mu_D) = 10698$ p = .017* Low PE: $E(\mu_D) = 1081$ p = .82	

Figure 16. Experimental condition combinations.

Results Relevant to Hypothesis Four: There is no support for this hypothesis. We expected to show via our omnibus test that Congruency Recall Advantage would be positively related to PE in the Black target/slow presentation condition. There was essentially zero correlation in this cell ($r = -.04$, $p = .42$, $n = 28$ subjects providing 266 combined total recalls). We had specifically predicted (in this Black/slow cell) that high PE scorers would recall a greater proportion of congruencies while low PE scorers would recall a greater proportion of incongruencies. Though high PE subjects did recall a slightly higher proportion of congruencies (.382 versus .357), this difference was not significant. Likewise, low PE subjects did recall a slightly higher proportion of incongruencies (.285 versus .281) — but again, this difference was not significant.

Results Relevant to Hypothesis Five: This hypothesis was simply intended to check our assumption that our White subjects' prejudicial expectations concerning Black targets would be irrelevant when the target was White. Fortunately, PE did not predict recall behavior in this cell (although something strange seems to be going on — see our final paragraph below). Correlations were low and non-significant ($r_{PE;SQE} = -.076$, $p = .40$, $r_{PE;CRA} = -.03$, $p = .46$, $n = 14$ subjects providing 133 combined total recalls). Likewise, using a median split (on PE) of subjects in this cell, the recall proportions compared in the Black target cells were non-significant here (p-values of .8874 and .3488 respectively with the low PE — probably brighter — subjects recalling more of both types). Of course, Figure 16 and the comments in the last paragraph above it (in the Hypothesis Three results) are pertinent in this section as well.

Finally, we should not pretend that the use of a White target represented an entirely null treatment. White subjects who've just had their prejudice measured — and are then presented with a White character named Carl Albert Mueller — are probably confused — and possibly a little hostile. It is not clear to us what is driving the recall proportions seen in this cell. It's likely we've just captured some peculiar sporadic recall behavior in this group of subjects. Removing four peculiar subjects from this cell leaves the Srull charts looking a little more consistent with our expectations (see Figure 19).

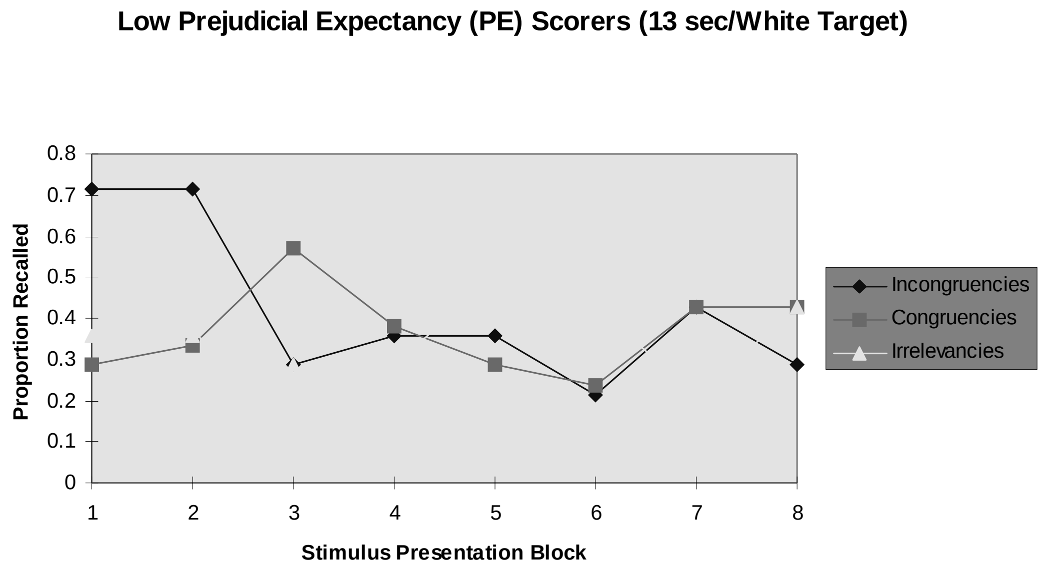

Figure 17. Srull Expectancy Emergence Chart

Summary of Results: In general, our data suggest that H-SAN person-memory schema effects do not manifest themselves when the initial impression relies on racial prejudice alone and the dimension of appraisal is academic aptitude or potential. In other words, the schema effects seen previously (when the initial impression has been provided entirely by the experimenter and the dimension of appraisal has typically been affability) may not generalize to our area of interest. Regrettably, we cannot identify

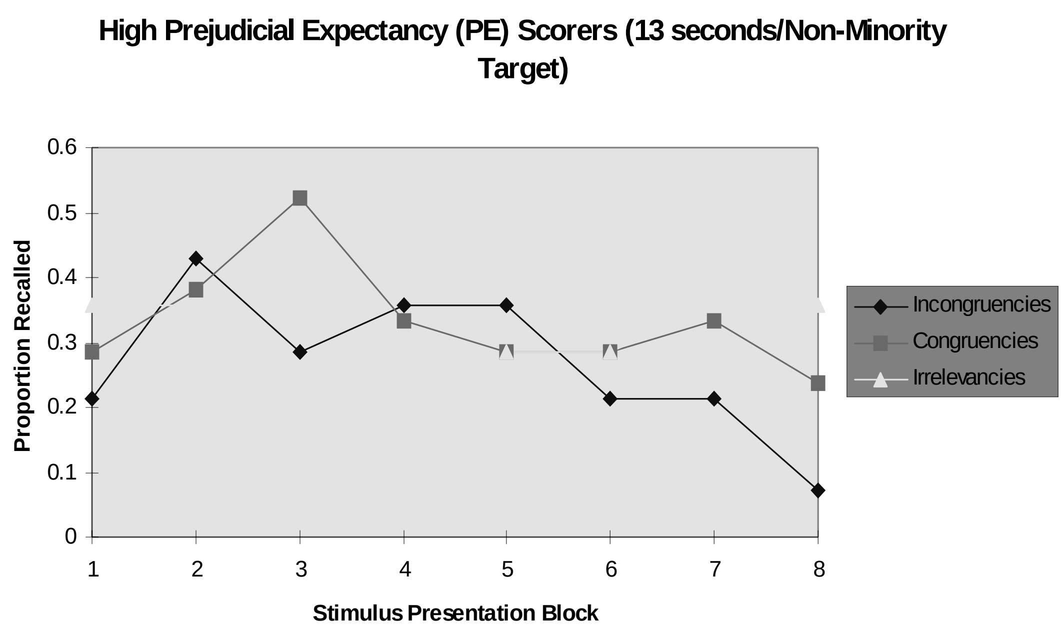

Figure 18. Srull Expectancy Emergence Chart

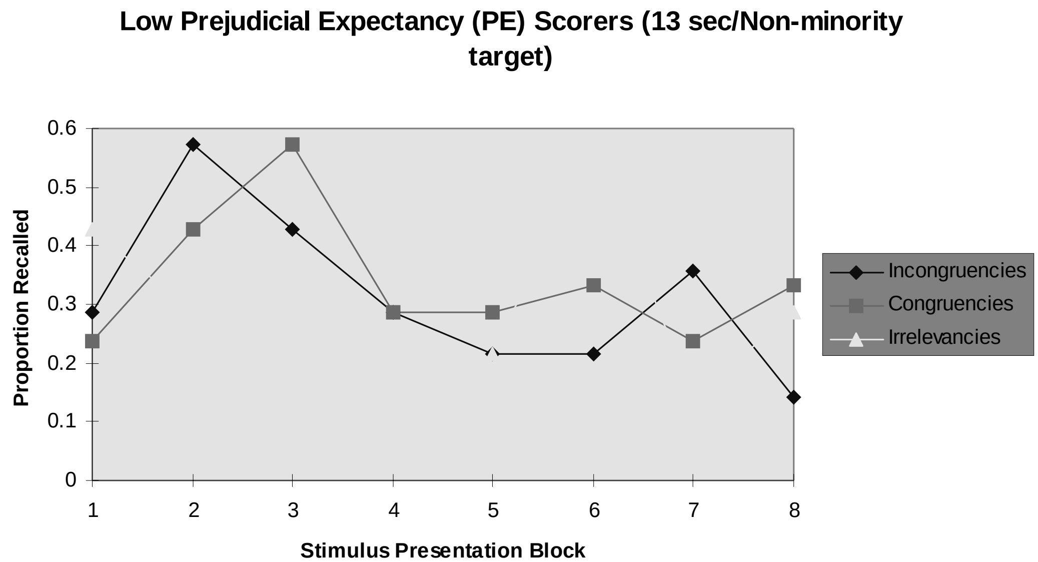

Figure 19. Srull Expectancy Emergence Chart

which feature of our design has diminished H-SAN related processes or its manifestations — the move from affability to performance potential *or* the move from experimenter-provided expectancies to the subjects' own enduring prejudicial expectations. Stangor and Ruble's (1989) work might suggest the latter was likely to be a problem.

Had we achieved the best of all possible outcomes, we could have suggested *why* well-entrenched expectancies do not manifest themselves as experimenter-provided expectancies do in these designs. Specifically, if our inter-recall interval data replicated Srull's (1985) findings, but the recall probabilities were more in-line with Stangor and Ruble's (1989) thinking, we would have suggested a post-encoding/mid- or post-retrieval filter. Ideally, we would hope to find that prejudice predicts schema effects in the inter-recall time intervals, but that recall probabilities were predicted by some interaction of prejudice and impression management. If such mechanisms are at work here, they are not revealed in our analysis. We suspect H-SAN effects may not generalize to the prejudice-performance appraisal setting.

We say the H-SAN *effects* may not generalize, because we believe it to be an overstatement to say H-SAN *mechanisms* may not generalize here. It is quite possible that long-term storage of person-tagged information still occurs in ways predicted by the H-SAN model when prejudiced raters appraise performance potential. These statements are justified by the marked appearance of some noticeable schema effects (mentioned at the beginning of this chapter and reflected in Figure 11).

Our design was not able to rule out the operation of some habitually or consciously-controlled mechanism in the retrieval or reporting process (the above-mentioned filter). This relates back to impression management – with which mean scores of our racial prejudice items were substantially correlated.