Evidence for a retrocausal effect in the human nervous system

Dean Radin & Edwin May
Boundary Institute

Four double-blind experiments tested the hypothesis that under certain conditions a future event may influence a past event. The past event was the time it took a person to identify a colored rectangle on a computer screen. The future event was a variant of the Stroop effect, which modulates reaction times due to cognitive-perceptual interference between pairs of colors and color-names. The "time-reversed interference" hypothesis postulates that there will be a positive correlation between past and randomly determined future events. After considering numerous alternative explanations for the observed results, we conclude that these experiments provide evidence for an unconscious time-reversed effect in the human nervous system (overall p = .001).

Introduction

Imagine that you are shown a page of color-names printed in various colored inks. You are asked to state the colors of the color-names as quickly as possible. It is well known that if the ink colors and color names are congruent (i.e., they match) then the time it takes you to complete this task will be shorter than if the colors and names are incongruent (mismatch). This differential effect, named after its discoverer, John Ridley Stroop (Stroop, 1935), is remarkably robust. The effect is attributed to cognitive interference between the perceptual processing of colors versus color names. Many variations of the Stroop effect have been studied and published in over 700 journal articles (MacLeod, 1991).

In the early 1980's, Holger Klintman (1983, 1984) was conducting a reaction time (RT) experiment based on a Stroop task. He was interested in improving the sensitivity of his measurements by calibrating each RT with a prior baseline RT. He asked people to first identify the color of a colored rectangle as quickly as possible, and then report whether a color-name (like the word "green") matched or mismatched the color of the rectangle. The initial color identification task was the baseline reaction time (RT1), and the second was a Stroop task (RT2).

To Klintman's surprise, he found that the variability in RT1 was larger than expected. Upon further investigation, he found that RT1 was faster when the color-patch and color-name matched, and slower when they mismatched. Because the match/mismatch condition was randomly determined by a hardware random number generator *after RT1 was already recorded*, the two conditions were equally probable, and the study followed conventional double-blind protocols, Klintman concluded that RT1 was dependent to some extent on RT2. He dubbed this anomaly "time reversed interference" (TRI). After

conducting five experiments, using different designs to provide conceptual replication, he concluded that the TRI effect was not a design or hardware artifact. The combined result for all five of his experiments was $p \approx 10^{-6}$.

A few years later, Gert Camfferman (1987) attempted a replication of Klintman's TRI effect. His participants saw equal numbers of trials with a color patch followed by a name (color-name task), or a name followed by a color (name-color), and the order was counterbalanced within participants. He found a significant difference in mean reaction times with the color-name task, but not with the name-color task. However, because he also discovered a positive correlation between RT1 and RT2, he concluded that the assumption that RT1 was independent of RT2 was wrong, and that the apparent TRI effects were more likely to be due to variations in general alertness than to some form of retro-causation. That is, if Alice had just quaffed a triple expresso, she would be faster on both RT1 and RT2 than Betty, who had just finished a bottle of wine. Camfferman's (1987) conclusions apparently dampened interest in Klintman's findings, because no further reports testing this hypothesis were reported.

However, while Camfferman's conclusion that RT1 and RT2 are related is undoubtedly correct, his assumption that Klintman's observations could be completely attributed to variations in alertness may have been premature. After all, Klintman hoped to exploit the dependence between RT1 and RT2 to form a more sensitive RT2, but in the process he discovered an unexpected differential effect that could not be explained by variations in alertness.

A decade after Klintman's puzzling observations, some new experiments have generated additional evidence for differential retro-causal effects in the human autonomic nervous system (e.g., Bechara, Damasio, Tranel & Damasio, 1997; Radin, 1997; Bierman and Radin, 1997, 1998; Radin, in press). Because of this new evidence, we decided to revisit Klintman's TRI effect, using a new design.

Method

Procedure

In Studies 1, 2, and 3, a participant (P) saw a red, blue, green or yellow rectangle in the center of an otherwise black computer screen. P was asked to type the letter r b g or y as quickly as possible to identify the color. Immediately after entering his or her choice, the colored rectangle disappeared and a color-word appeared. P now typed y for yes, meaning the color word *matched* the preceding color patch, or n for no, meaning it *mismatched*. One trial in these experiments consisted of this pair of RTs.

This procedure differed from Klintman's design in two ways. First, Klintman used a voice-key to stop the timing clock; we asked people to type their response. Thus, instead of a single baseline RT to examine, we had four baseline RTs due to the

different amounts of finger-movement time it takes to indicate four colors by typing. Second, Klintman used a maximum latency of 850 milliseconds between RT1 and RT2; we set a 5 second maximum latency on both RT1 and RT2 (because typing responses are typically longer than vocal responses).

One session of 20 such trials generally took about five minutes to complete, and the task was easy to explain and perform. If P's typed identification of the initial color-patch was incorrect, the program displayed the message "Whoops, try again," and that trial was repeated with newly re-randomized color-patch and match/mismatch conditions.

Study 4 differed from the first three in that P was asked to type the first letter of an initial color-patch (as before), but then typed the *color* of the following color-name, rather than y or n. This task was closer to the classical form of the Stroop task and was therefore expected to provide a stronger differential effect in RT2.

Software

Study 1 was conducted using a DOS program written in Microsoft QuickBasic 4.5.[1] This program provided millisecond reaction time accuracy through the use of special assembly language timing routines written for use in reaction time tasks (Graves & Bradley, 1987, 1988, 1991). Studies 2 through 4 were Windows applications written in Microsoft C++ (version 5.0) [2], following the design of the original QuickBasic program. These programs provided timing accuracy through the use of a high-resolution performance counter available through C++. In addition, because Windows has processes running in the background, when the C++ experimental program was launched it immediately raised its processing status to the "real-time" priority level. This reduced the possibility that background processes might interfere with the experimental timing.

The random selection of match/mismatch conditions in Studies 1-3 were controlled by a pseudorandom number generator (PRNG) algorithm. In Study 1 the QuickBasic "RND" function was used, and in Studies 2 and 3 the C++ rand() function was used. For the first half of Study 1, the PRNG was reseeded once at the beginning of the program, and for the second half the PRNG was reseeded at each successive button press using the current system clock time.

In Studies 2 and 3, the PRNG was reseeded at each button press using the current system clock time ($1/60^{th}$ second clock rate). The first random number generated after re-seeding was used to form a random number N from 1 to 1000, and then N additional

[1] By the first author.

[2] By Steve Rubin and the first author.

random numbers were immediately generated. The N + 1 random number was then taken modulo 2 to select whether the upcoming condition would be a match or mismatch. Recall that this match/mismatch random decision occurred *after* the first reaction time was already recorded.

In Study 4, an electronic noise-based RNG was used for all random decisions. In all Studies, the initial color-patch color was uniformly selected from the four color possibilities, and the subsequent match/mismatch condition was a random binary variable. If the condition turned out to be a mismatch, then the mismatched color name was selected uniformly at random from the three remaining color-name possibilities.

Data and Analysis

Reaction-time data were normalized per participant P into standard normal deviate scores. The normalization was conducted to take into account different individual baseline reaction times, and it also allowed data to be pooled across Ps. For example, for each RT1 measurement, Z_{RT1} was formed as

$$Z_{RT1} = (x-\mu)/\sigma,$$

where Z was distributed as N(0,1), x was RT1 in a given trial, μ was the mean of all RT1 values for a given P in a given session over both match/mismatch conditions, and σ was the standard deviation of all RT1 values in that same session, again over both conditions. The same normalization was performed independently for RT2, using μ and σ from the RT2 values to create a Z_{RT2} score for each trial.

Then two differential measures were formed:

$$\Delta Z_{RT1} = (Z_{RT1.Mismatch} - Z_{RT1.Match})/\sqrt{2}, \text{ and}$$

$$\Delta Z_{RT2} = (Z_{RT2.Mismatch} - Z_{RT2.Match})/\sqrt{2}, \text{ where}$$

$Z_{RT1.Mismatch} = \sum Z_{RT1.Mismatch} / \sqrt{[N_{Mismatch}]}$, summed over all $RT1._{Mismatch}$ trials, $N_{Mismatch}$ = count of all trials in the Mismatch condition, and similarly for the other terms in the equations.

To confirm that the probabilities associated with the ΔZ measures (based on the normal curve) were valid, randomized permutation analysis was used (Diaconis & Efron, 1983; Efron & Tibshirani, 1991; Good, 1994; Hjorth, 1994). See the Appendix for a description of this method and a comparison of the theoretical probabilities associated with ΔZ_{RT1} and ΔZ_{RT2} vs. the probabilities resulting from a randomized permutation analysis.

Hypothesis

The time-reversed interference (TRI) hypothesis states that modulation of RT2 will be retroactively mimicked in RT1. RT2 is modulated in the present experiments through the application of cognitive-perceptual interference. Formally, TRI predicts that ΔZ_{RT1} will be positively correlated with ΔZ_{RT2}. From the classic Stroop effect we might expect $\Delta Z_{RT2} > 0$, however because of the different motor-movement times associated with typing different keys in this experiment, we did not predict that each initial color would independently result in a Stroop effect, or even that the sign of ΔZ_{RT2} would necessarily be positive for each color. To further assist in clarifying the nature of this hypothesis, refer to Figures 1, 2 and 3 and the associated captions.

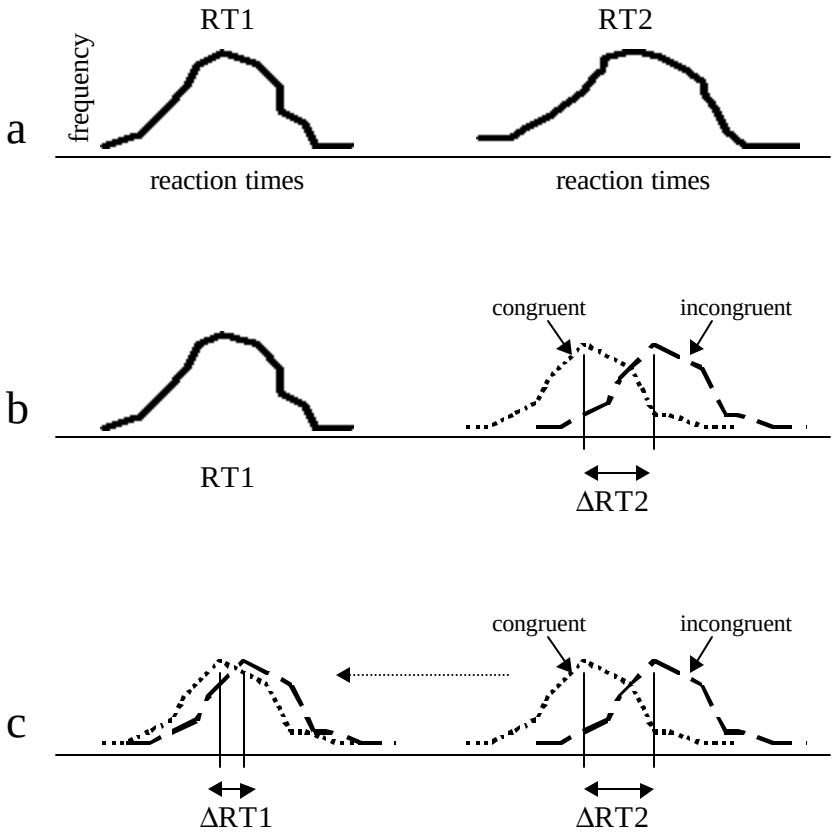

Figure 1. Illustration of the time-reversed interference (TRI) hypothesis. (a) Representative frequency distributions of reaction times for the initial color-identification task, labeled RT1, and for the following match / mismatch decision task, labeled RT2. (b) The second task manipulates reaction times with cognitive-perceptual interference, thus the RT2 distribution consists of two sub-distributions, one for congruent responses and the other for incongruent responses. The mean difference between these two distributions is called ΔRT2. (c) The TRI hypothesis postulates that (1) cognitive-perceptual interference at RT2 results in ΔRT2, that (2) this interference sometimes propagates "backwards in time" to modulate the RT1 distribution, and that (3) this will cause ΔRT1 to mimic ΔRT2.

This postulate is formally tested by examining the correlation between ΔRT2 and ΔRT1.

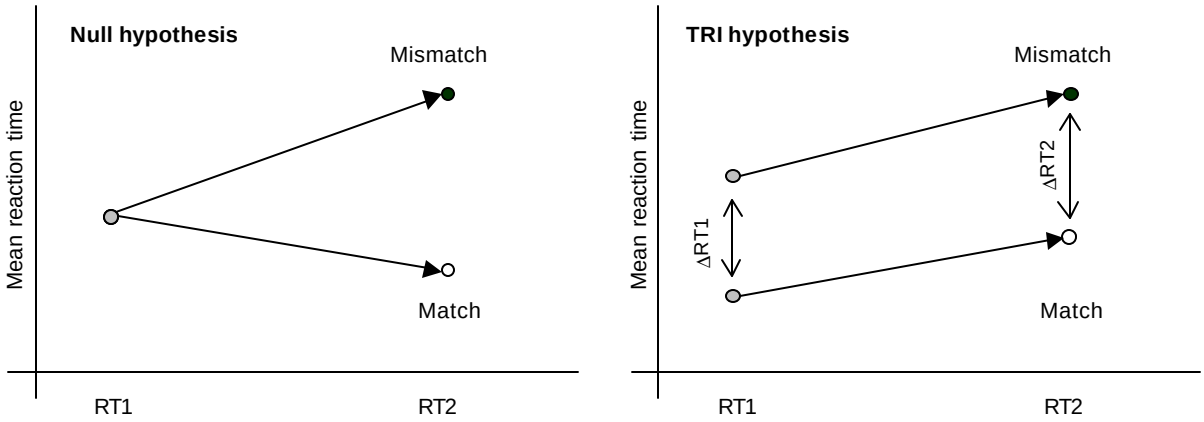

Figure 2. The null hypothesis predicts that mean reaction times at RT1 will not depend on the match / mismatch condition at RT2 because the condition is selected in the future. Thus, on average, the difference between RT1(match) vs. RT1(mismatch) will be zero, or $\Delta RT1 = 0$. In contrast, the cognitive-perceptual interference introduced by the match / mismatch task at RT2 will modulate ΔRT2, thus on average $\Delta RT2 \neq 0$. If this interference has a time-reversed influence, then we would predict a positive correlation between ΔRT2 and ΔRT1.

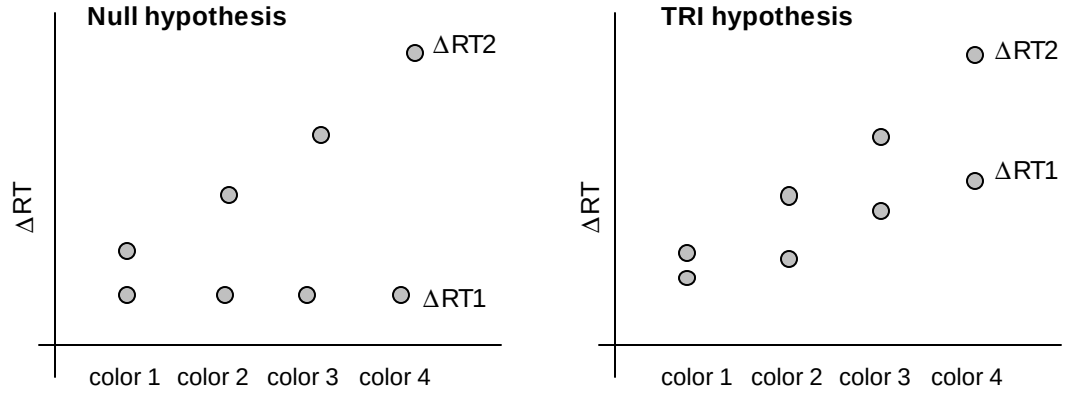

Figure 3. The null hypothesis predicts a zero correlation between the difference in mean reaction times, ΔRT1 and ΔRT2, across the four colors used in the experiment. Note that the Stroop effect at RT2 is not expected to produce the same results for each of the four colors. The TRI hypothesis predicts a positive correlation between ΔRT1 and ΔRT2, under the assumption that the interference that caused the ΔRT2 difference will leak backwards in time and affect ΔRT1 in a similar way.

Results

Study 1

Study 1 was a pilot test conducted by one participant who contributed 500 trials,[3] 254 of which were randomly assigned to the mismatch condition and 246 to the match condition. Figure 4 shows the results of Z_{RT1} and Z_{RT2} for match / mismatch conditions, for the color green. The difference between these values (i.e., ΔZ_{RT1} and ΔZ_{RT2}) mimics a similar pattern of results observed at RT1, as predicted by the TRI hypothesis.

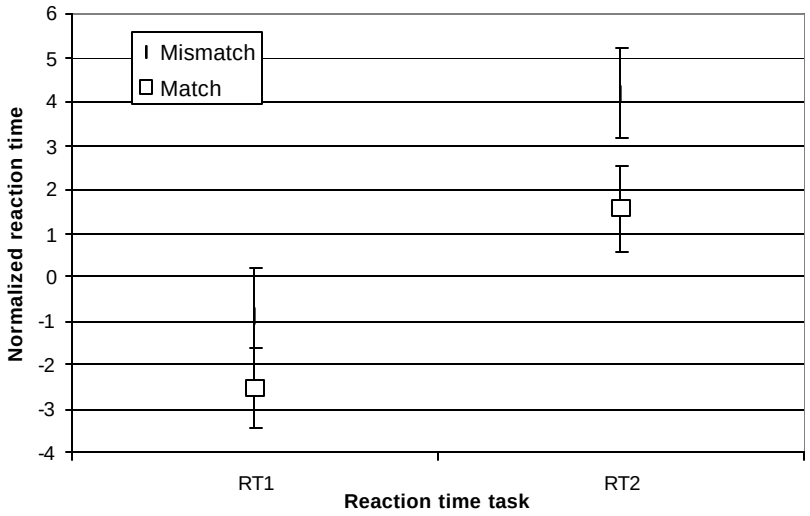

Figure 4. Normalized reactions times for RT1 and RT2, in the match and mismatch conditions, with one standard error bars.

Figure 5 graphs ΔZ_{RT1} and ΔZ_{RT2} for all four colors; Table 1 lists the number of trials, per color, in the two conditions. The correlation between ΔZ_{RT1} and ΔZ_{RT2} for four pairs of points results in a positive, but non-significant correlation. r = 0.309, p = 0.345, one-tail.

	Mismatch	Match
Red	72	56
Yellow	68	70
Blue	60	63
Green	54	56

Table 1. Number of trials in each color, by condition, in Study 1.

[3] All participants in these experiments are anonymous, as required by Interval's Human Subjects Guidelines.

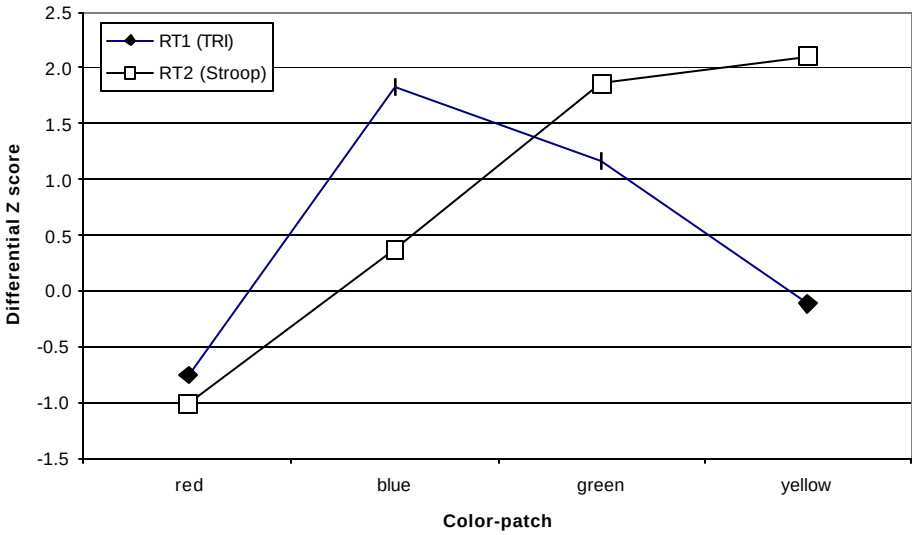

Figure 5. Mean values of ΔZRT1 and ΔZRT2, by color. The correlation between these curves is r = 0.309 (p = 0.345, one-tail).

Study 2

Study 2 was conducted as a pilot test of a C++/Windows 95 version of this experiment. The new program was designed to explore whether the small positive correlation observed in the first study was due solely to chance, or to possible programming artifacts. Ten participants contributed a total of 1,852 usable trials.[4]

As shown in Figure 6, the correlation between ΔZ_{RT1} and ΔZ_{RT2} by initial color-patch was in the predicted direction, and was nearly significant, r = 0.842 (p = 0.079, one-tail). Table 2 shows the number of trials per color, by condition.

	Mismatch	Match
Blue	221	254
Green	237	244
Red	233	218
Yellow	216	229

Table 2. Number of trials in each color, by condition, in Study 2.

[4] A few trials with reaction times greater than 5 seconds were dropped from the analysis. Participants contributed from 1 to 50 sessions of 20 trials each.

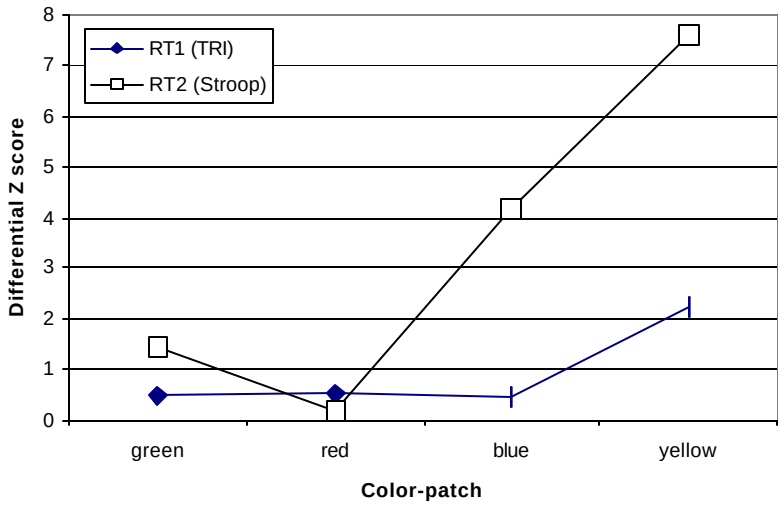

Figure 6. Mean values of ΔZRT1 and ΔZRT2, by color. The correlation between these curves is r = 0.842 (p = 0.079, one-tail).

Study 3

Study 3 was a formal replication of Study 2. The experiment was advertised on Interval Research Corporation's internal web site. Anyone who wanted to participate in the experiment could anonymously download the test program and conduct the experiment at their desk on their own PC, at will. The program provided instructions on how to conduct the test, it presented the user with six practice trials, then it ran 20 formal trials. Only data from formal trials were recorded.

Each person was asked to provide three sessions of 20 formal trials each. However, because all data in this study were collected anonymously, there was no way to know if all participants completed three sessions. When each person finished a session, the program sent the data over the internal network to a remote file. To ensure data integrity, a checksum was secretly added to each data record so the entire dataset could later be verified.[5] After two weeks, the experiment was again advertised on the internal web site, and some additional data were recorded.

At the end of a month,[6] a total of 111 sessions of 20 trials and one session of 30 trials were recorded, totaling 2,270 trials, of which 1,160 were in the match condition and 1,110 were in the mismatch condition.[7] Of these trials, 47 had reaction times greater than 5 seconds in either RT1 or RT2; these trials were eliminated from further analysis.

[5] There was no indication that any data collected in this study had been tampered with.

[6] The experiment ran from December 18, 1998 to January 17, 1999.

[7] One person contributed a single session of 30 trials, as part of a demonstration by the first author.

This left 2,223 trials with usable reaction times (1,138 match and 1,085 mismatch). Results, shown in Figures 7 and 8, indicate a significantly positive correlation between ΔZ_{RT1} and ΔZ_{RT2}, separated by initial color-patch, r = 0. 989 (p = 0. 005, one-tail). Table 3 shows the number of trials per color, by condition.

	mismatch	match
blue	286	277
Green	267	292
Red	279	277
Yellow	253	292

Table 3. Number of trials in each color, by condition, in Study 3.

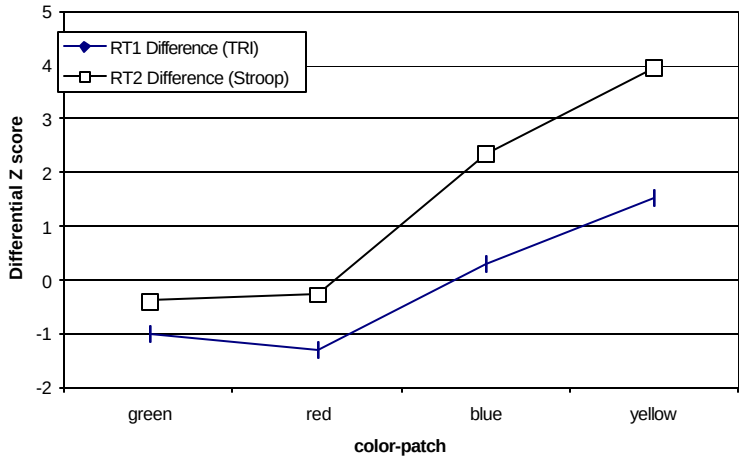

Figure 7. Mean values of ΔZ_{RT1} and ΔZ_{RT2}, by color. The correlation between these curves is r = 0.989 (p = 0.005, one-tail).

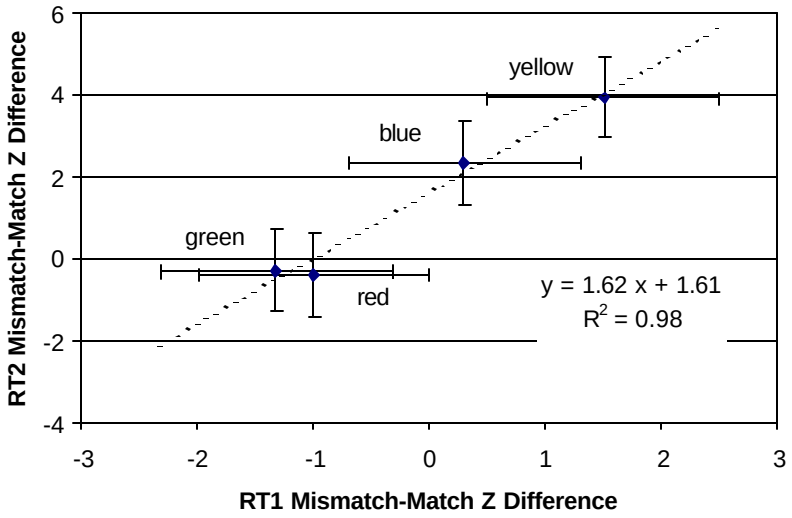

Figure 8. Same data as in Figure 7, showing the linear regression line, r = 0.989, p = 0.005, one-tail).

Study 4

Study 4 was designed to explore what would happen to the correlation of interest when very large ΔZ_{RT2} values were generated. To do this, rather than ask participants to decide if a color-name matched or mismatched a initial color-patch (as in the first three experiments), they were asked to type the first letter of the *color* that the color-name was printed in. This is similar to the original task reported by Stroop in 1935.

Two participants produced a total of 720 trials, 366 mismatches and 354 matches; Table 4 shows the number of trials per color, by condition. As shown in Figure 9, the correlation between ΔZ_{RT1} and ΔZ_{RT2} by initial color-patch was significantly in the predicted direction, r = 0. 926 (p = 0. 037, one-tail). The large ΔZ_{RT2} values did not seem to significantly enhance or to reduce the hypothesized TRI effect.

	mismatch	match
blue	100	93
green	94	85
red	85	92
yellow	87	84

Table 4. Number of trials in each color, by condition, in Study 4.

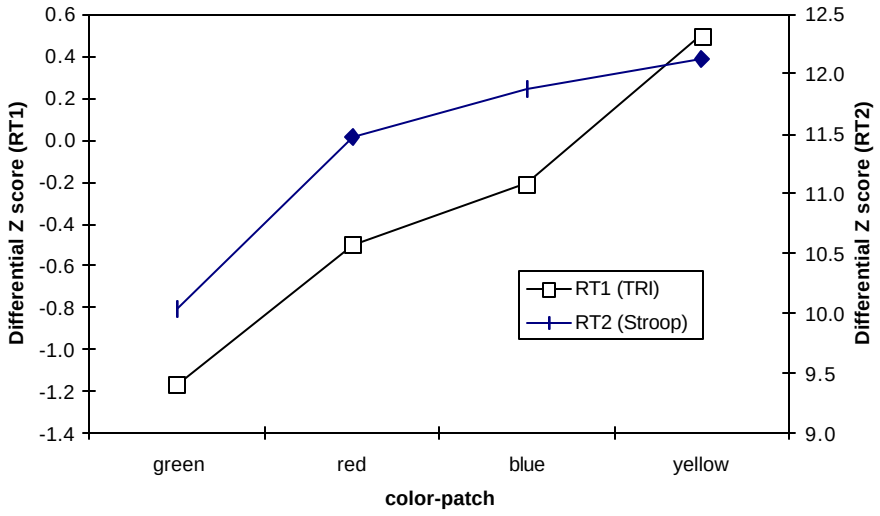

Figure 9. Mean values of ΔZRT1 and ΔZRT2, by color. The correlation between these curves is r = 0. 926 (p = 0. 037, one-tail).

Discussion

The time-reversed interference (TRI) hypothesis predicts positive correlations between ΔZ_{RT1} and ΔZ_{RT2}. Positive correlations were observed in all four experiments, two of which were significant and one was marginally significant. Overall, these experiments allow us to reject the null hypothesis (Stouffer Z = -3.08, p = 0.001).[8]

Are these positive correlations due to the hypothesized time-reversed interference, or to a more prosaic causal link, RT1→RT2? This latter correlation reflects an expected forward-time dependency between two reaction times collected close in time, and to fluctuations in alertness as proposed by Camfferman (1987). The answer should be no, because the experimental protocol is designed to scramble any causal dependencies between RT1 in the present and RT2 in the future. However, if the arousal factor happens to match the future match / mismatch conditions even slightly on average, then it is plausible that forward-time dependencies may statistically compound over many repeated trials and cause the observed positive correlations between ΔZ_{RT1} and ΔZ_{RT2}. To explore this in more detail, let us examine the evidence for what we'll call the "alertness hypothesis."

Alertness Hypothesis

A single dataset was created from the combined normalized data collected in Studies 2 and 3. These two datasets were pooled because they involved the same task and used the same Windows 95 program to collect the data. In examining this data, it was apparent that a small number of trials resulted in long RT2 latencies. A second dataset was therefore created in which a small number of outliers, defined as trials where RT2 > 2 seconds, were removed. This new dataset contained about 98% of the original dataset.

Figure 10 shows the results of correlating Z_{RT1} vs. Z_{RT2} for the entire database and for the outlier-eliminated subset, along with one standard error bars.[9] There was an overall correlation of approximately r = 0.14 in both datasets. The other columns in Figure 10 indicate that the $Z_{RT1} \leftrightarrow Z_{RT2}$ correlation was somewhat stronger for mismatches than for matches, suggesting that the $Z_{RT1} \leftrightarrow Z_{RT2}$ relationship interacted with cognitive-perceptual interference.

[8] The Stouffer Z estimate is based on transforming the correlation p-value from each study into a one-tailed Z, then combining those Z scores.

[9] When using data pooled across subjects, it is important to correlate normalized reaction times, Z_{RT1} vs. Z_{RT2}, - rather than RT1 vs. RT2, because otherwise an artifactually high correlation will be produced due to large between-subject differences in baseline reaction times.

These correlations imply that the alertness factor was *not* completely nullified in studies 2 and 3. Thus, it is reasonable to ask whether this small correlation adequately explains the results of these experiments.

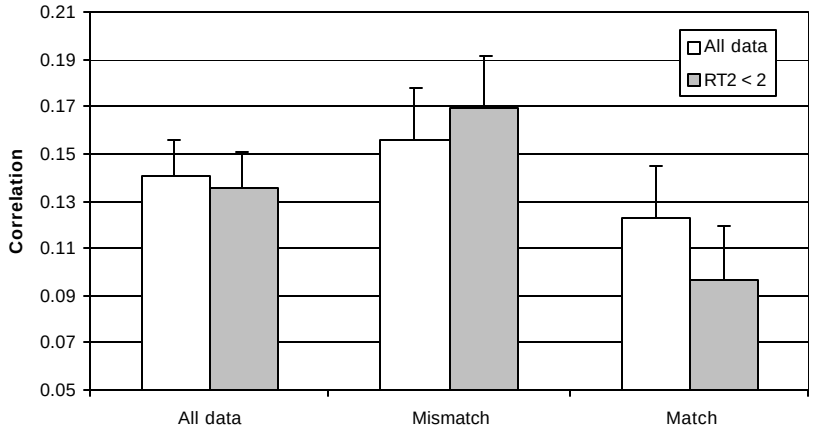

Figure 10. Correlation between Z_{RT1} and Z_{RT2}, for all data and for a subset dataset excluding all trials in which RT2 > 2 seconds.

To study this question, first we assumed that variations in alertness carry over from RT1 to RT2. This means that when RT1 was slightly faster, then RT2 was slightly faster as well, and vice versa. We further assumed that this positive, forward-time dependency ultimately resulted in the positive ΔZ_{RT1} vs. ΔZ_{RT2} correlations observed in the four studies reported here.

To test this conjecture, we selected out those 20-trial sessions from Study 3 where the $Z_{RT1} \rightarrow Z_{RT2}$ correlation was *negative*. If the above postulate is correct, and the RT1→RT2 relationship does indeed drive the ΔZ_{RT1} vs. ΔZ_{RT2} outcome, then this subset should result in a *negative* $\Delta Z_{RT1} \rightarrow \Delta Z_{RT2}$ correlation.

The average per-session $Z_{RT1} \rightarrow Z_{RT2}$ correlation for all data in Study 3 was r = 0.20. Of all 120 20-trial sessions in Study 3, per-session correlations ranged from r = -0.46 to r = 0.74. We selected out those sessions with the most negative correlations (ranging from r = -0.10 to –0.46), then re-calculated the final $\Delta Z_{RT1} \rightarrow \Delta Z_{RT2}$ correlation based on this subset of trials. The result, as shown in Table 5, was r = 0.77, p = 0.12. While not independently significant, this *positive* correlation counters the postulate that a forward-causal relationship drives the results of these experiments. Even with quite strong negative relationships between RT1 and RT2, the final $\Delta Z_{RT1} \rightarrow \Delta Z_{RT2}$ correlation was still positive.

Color	ΔZ_{RT1}	ΔZ_{RT2}
Blue	-1.74	-0.29

Red	0.33	-0.14
Green	0.50	1.00
Yellow	1.08	1.14

Table 5. The correlation between ΔZ_{RT1} and ΔZ_{RT2} for the most negative sessions was r = 0.77, t (2df) = 1.68, p = 0.12.

Cross-experimental consistency

If time-reversed interference reflects a genuine, unconscious retrocausal effect, we might expect it to be consistent across conceptually identical experiments.[10] To test this, the combined statistical results for ΔZ_{RT1} and ΔZ_{RT2} was determined per color using the Stouffer Z method. The results, shown in Table 5, confirms that as ΔZ_{RT2} increased, ΔZ_{RT1} also increased. Another way of demonstrating the same relationship is to plot the correlation between ΔZ_{RT1} and ΔZ_{RT2}, per color, over the first three experiments. The results, shown in Figure 11, confirm the predicted correlation, r = 0.677, p = 0.008, one-tailed (10 df).

INITIAL COLOR	ΔZ_{RT1}	ΔZ_{RT2}	p(RT1)	p(RT2)
Red	-0.862	-0.475	0.806	0.683
Green	0.494	1.680	0.311	0.046
Blue	1.541	4.049	0.062	0.00003
Yellow	2.169	7.952	0.015	<< .001

Table 5. Combined results (Stouffer Z) for each color separately, from the first three studies, with associated one-tailed p-values.

[10] That is, unconscious tasks are more likely to bypass the psychological defense mechanisms that often interfere with conscious performance. Results from Study 4 were not included in this analysis because it involved a different task.

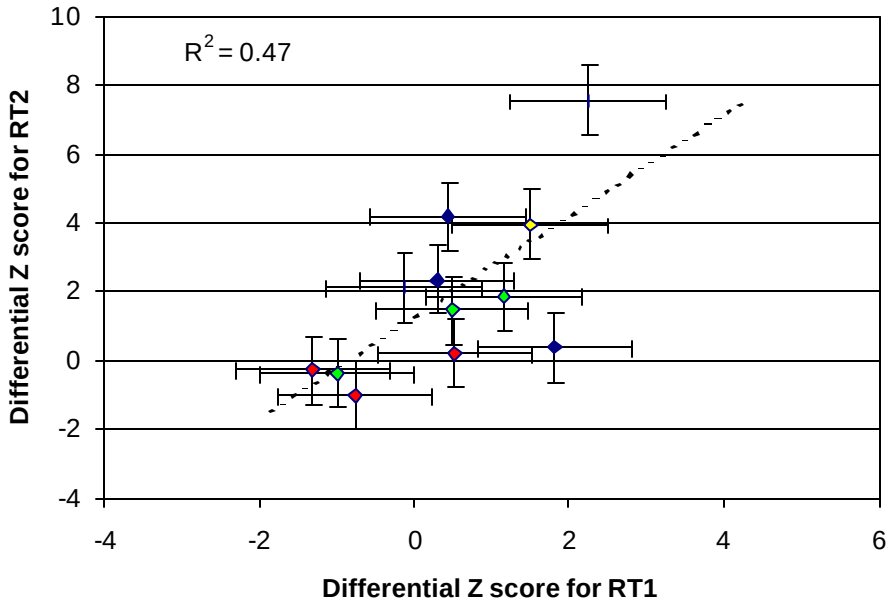

Figure 11. Correlation between ΔZ_{RT1} and ΔZ_{RT2} by color, for first three experiments, N = 12, r = 0.677, p = 0.008 (one-tailed).

Alternative Explanations

If leakage of forward-time dependencies between RT1 and RT2 is not a viable explanation, what else might account for the present results without entertaining exotic ideas like retrocausal influences? There are six primary alternatives: (1) sensory cues, (2) statistical cues, (3) systematic timing problems, (5) selective data reporting, and (6) optional stopping.

Sensory cues

If participant P obtained a sensory cue about the upcoming match / mismatch condition, this might bias P to respond differently before the two conditions, and hence to possibly generate results similar to those observed here. Such cues might occur as a result of say, different sounds being generated by the computer's hard disk when the next trial was going to be a match.

We believe that sensory cueing is an unlikely explanation for the obtained results because (a) the experiments were conducted under double-blind conditions, with neither experimenter nor subject knowing the condition of each subsequent trial, (b) no feedback was provided about reaction times, and (c) the future match / mismatch conditions were not generated until *after* the first reaction time was already completed. Thus, at the time of the first reaction time there was no information available to provide any form of sensory cueing.

Statistical hints

Like sensory cues, statistical hints about the upcoming condition might unconsciously clue a participant into responding differently from one trial to the next. To check for non-random patterns that might have provided statistical hints, the sequence lengths between successive "match" conditions were examined in the combined data from Studies 2 and 3. The histogram in Figure 12 shows that the observed sequence lengths closely matched the periods theoretically expected from a good random sequence.

Recall that all random decisions in Studies 1-3 were based on the output of pseudorandom generators that were re-seeded after each button press. In Study 4 a hardware (electronic noised-based) random number generator was used to make all random decisions. The fact that similar results were observed in these studies suggests that the precise form of randomization was not an important factor in the observed results.

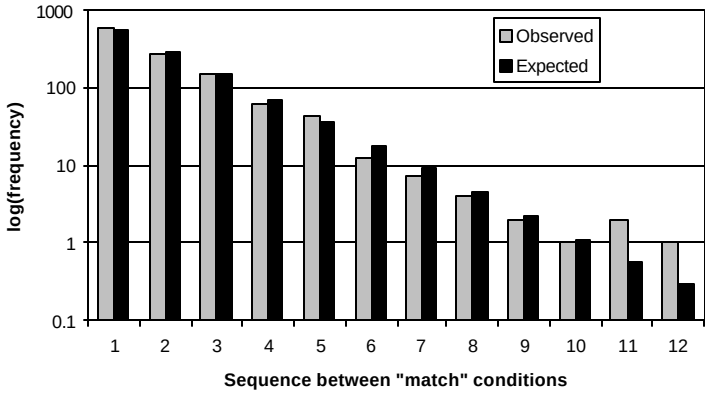

Figure 12. Histogram of log(frequency) vs. sequence lengths for observed and expected "match" conditions.

Timing problems

The essential data in these experiments were timing measurements. If there were any small timing errors that were *differentially* preferential towards one or another condition, these may have accumulated over the course of thousands of trials to produce spurious results. As previously mentioned, to offset this possibility, the timing routines in Studies 2, 3 and 4 relied on a hardware clock, not on software loops. In addition, to avoid other delays due to background processes running in the computer, the experimental program's processing priority was raised to the highest level.

To further check timing accuracy, a routine was added to the (Windows) experimental code to allow the program to act as a stopwatch. A test was then conducted in which the experimental program was run in the background while other programs were used.

The program's stopwatch was halted after the passage of 2,264.26 seconds. The external stopwatch was halted at the same time (within human reaction time). It had recorded 37minutes, 44.57 seconds. The difference is 310 milliseconds, which given the total length of time, means that the program's clock ran about 140 microseconds per second slower than the stopwatch. This is at least two orders of magnitude smaller than the effects observed in the present experiments.

In any case, it seems unlikely that timing artifacts could have produced the results observed in these studies because the measurements of primary interest (i.e., RT1) were recorded *before* the future match/mismatch condition was generated. Say, for example, that the computer was "distracted" in one trial by a background task. This might result in slower reaction time measurements for both RT1 and RT2 in that trial. If these distractions occurred often, they might eventually form a spurious correlation between RT1 and RT2. But for such timing artifacts to explain the present results, we would have to postulate that somehow the computer became systematically distracted only before the randomly selected match condition, and not before the mismatch condition (or vice versa). We could not find any evidence suggesting that such a differential timing artifact in fact occurred, or even how it might have occurred in principle.

Selective data reporting

To avoid the problem of statistical inflation associated with reporting only the "good" data, all data collected in these studies are reported here. No other pilot or formal studies were conducted.

Optional stopping

In Studies 2, 3 and 4, participants were allowed to conduct as many sessions as they wished in blocks of 20 trials. However, by design all of these studies provided no performance feedback, either on a per-trial or a per-session basis. Thus, participants in all of these studies could not take advantage of optional stopping strategies. Study 1 was pre-planned at 500 trials, and Study 3 was a formal test pre-planned for one month of data collection. Studies 2 and 4 were pilot tests without a pre-planned number of trials.

Conclusion

Common sense maintains that causality can flow only from past to future. But in spite of common sense, a growing number of scientists – mainly physicists – are seriously considering the possibility of genuine time-reversed effects. The scholarly literature contains over a hundred articles proposing ways in which retrocausality may be consistent with known physics. These articles are not found in fringe sources, but in

mainstream journals like *Physical Review*, *Science*, *Foundations of Physics*, and *American Scientist*.[11]

To test the possibility of detecting retrocausal effects in human behavior, four double-blind experiments were conducted. The experiments tested a time-reversed interference hypothesis using reaction times to simple tasks. The predicted relationship, a positive correlation between the strength of differential reaction times in the future versus the past, was observed in all four studies, and significantly so in two of the studies. The overall probability for these results was p = .001. We conclude that these studies provide support for the existence of a retrocausal effect.

References

Bechara, A., Damasio, H., Tranel, D., Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, **275**, 28 February 97, 1293-1295.

Bierman, D. & Radin, D. I. (1997). Anomalous anticipatory response on randomized future conditions. *Perceptual and Motor Skills*, **84**, 689-690.

Bierman, D. & Radin, D. I. (1998). Non conscious processes and intuition: Is there an anomalous component? *Consciousness Research Abstracts, Tuscon III: Towards a science of consciousness*. University of Arizona, Tuscon, April 1998.

Bierman, D. & Radin, D. I. (in press). Anomalous unconscious emotional responses: Evidence for a reversal of the arrow of time. To appear in *Tuscon III: Towards a science of consciousness*, MIT Press.

Camfferman, G. (1987). Time reversed interference: A replication study. *European Journal of Parapsychology*, **7**, 13-31,

Diaconis, P. & Efron, B. (1983, May). Computer-intensive methods in statistics. *Scientific American*, **48**, 116-130.

Efron, B. & Tibshirani, R. (1991). Statistical data analysis in the computer age. *Science*, **253**, 390-395.

Good, P. (1994). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. New York: Springer-Verlag.

Graves, R., & Bradley, R. (1987). Millisecond interval timer and auditory reaction time programs for the IBM PC. *Behavior Research Methods, Instruments, & Computers*, **19**, 30-35.

Graves, R., & Bradley, R. (1988). More on millisecond timing and tachistoscope applications for the IBM PC. *Behavior Research Methods, Instruments, & Computers*, **20**, 408-412.

Graves, R., & Bradley, R. (1991). Millisecond timing on the IBM PC/XT/AT PS/2: A review of the options and corrections for the Graves and Bradley algorithm. *Behavior Research Methods, Instruments, & Computers*, **23**, xxx-xxx.

---

[11] E.g., Tipler, Frank J.(1974). Rotating cylinders and the possibility of global causality violations, *Physical Review D*, Vol.9, No.8, 2203-220; Travis, John (1992). Could a pair of cosmic strings open a route into the past? *Science*, Vol.256, pp.179-180; Reitdijk, C. W. (1987). Retroactive effects from measurements, *Foundations of Physics*, Vol.17, 297-311; Elsasser, W. M. (1969). Acausal phenomena in physics and biology, *American Scientist*, Vol.57, pp.502-516.

Hjorth, J. S. U. (1994). *Computer intensive statistical methods: Validation model selection and bootstrap*. New York: Chapman & Hall.

Klintman, H. (1983). Is there a paranormal (precognitive) influence in certain types of perceptual sequences? Part I. *European Journal of Parapsychology*, **5**, 19-49.

Klintman, H. (1984). Is there a paranormal (precognitive) influence in certain types of perceptual sequences? Part II. *European Journal of Parapsychology*, **5**, 125-140.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, **109**, 163-203.

Radin, D. I. (1997). Unconscious perception of future emotions: An experiment in presentiment. *Journal of Scientific Exploration*, **11** (2), 163-180.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, **18**, 643-622.

Appendix

Randomized permutation analysis (RPA) was used to confirm that the ΔZ_{RT1} and ΔZ_{RT2} values were valid statistics for measuring the difference between match vs. mismatch normalized reaction times.

In descriptive terms, the RPA method first forms the difference $\delta_{RT1} = [Z_{RT1.Mismatch.Mean} - Z_{RT1.Match..Mean}]$, where $Z_{RT1.Mismatch.Mean}$ is the normalized mean reaction time for RT1 in the mismatch condition. The value δ_{RT1} is then compared to a distribution of δ_{RT1}'s formed by randomly scrambling the actual match / mismatch condition assignments, recalculating the resulting δ_{RT1}'s, and repeating this process 1,000 times. If the original δ_{RT1} was larger than, say, 900 of the randomly permuted δ_{RT1}'s, then the associated p-value for δ_{RT1} would be (1000-900) / 1000 = 0.1. This p-value can then be turned into a ΔZ_{RT1} equivalent through the use of an inverse normal transform.

Table 6 shows the theoretical probabilities associated with ΔZ_{RT1} and ΔZ_{RT2} values for each initial color-patch, based on the combined data from Studies 2 and 3, in the column labeled P(Z Diff). The adjacent column shows probability estimates for the same differences, after using RPA with 1,000 permutations. Figure 13 shows the correlation between these two columns, confirming that the theoretical estimates of the probabilities are valid.

TASK	COLOR	N MISS	N MATCH	AVERAGE MISS	AVERAGE MATCH	Z MISS	Z MATCH	Z DIFF	P(Z DIFF)	P(RPA)
RT1	Red	494	524	-0.11	-0.10	-2.35	-2.21	-0.10	0.541	0.585
RT1	Blue	513	482	-0.21	-0.20	-4.73	-4.37	-0.25	0.600	0.513
RT1	Green	456	506	0.34	0.14	7.21	3.25	2.80	0.003	0.006
RT1	Yellow	497	521	0.10	0.05	2.13	1.22	0.64	0.260	0.256
RT2	Red	494	524	0.18	0.12	4.09	2.76	0.94	0.174	0.139
RT2	Blue	513	482	-0.08	-0.10	-1.75	-2.15	0.28	0.390	0.369
RT2	Green	456	506	-0.18	-0.68	-3.93	-15.23	7.99	0.000	0.000
RT2	Yellow	497	521	0.49	0.19	10.93	4.35	4.65	0.000	0.000

Table 6. Relationship between RPA estimated p-values (1,000 iterations) and theoretical p-values based on normal curve. N Miss and Match are the number of trials, Average Miss and Match are the mean normalized reaction time scores, Z Miss and Match are the means times $\sqrt{(N)}$, Z Diff is ΔZ_{RT1} and ΔZ_{RT2} as defined earlier, and p(Z Diff) is the one-tailed probability associated with ΔZ. The last column is the probability resulting from the randomized permutation analysis.

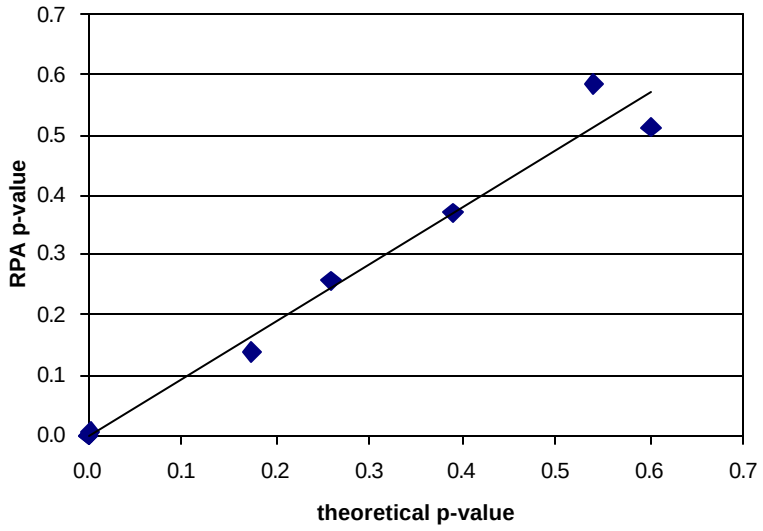

Figure 13. Relationship between RPA estimated p-values (1,000 iterations) versus theoretical p-values based on normal curve.