Similarity and Categorization: The Reversed Association Test

Takashi Yamauchi (tya@psyc.tamu.edu)
Department of Psychology, Mail Stop 4235
Texas A&M University College Station, TX 77843 USA

Abstract

The reversed association theory (Dunn & Kirsner, 1988) provides a powerful procedure for studying the link between cognitive processes and task performance. It helps find whether two behavioral tasks involve the same or different cognitive processes. However, this theory has not been fully utilized. Finding reversed association requires a large, single study consisting of at least six independent manipulations. Furthermore, the statistical procedure to verify reversed association has not been fully developed. This article presents practical solutions for these problems by investigating the recent controversy over categorization and similarity judgments. First, a contrast analysis is illustrated in a case study to statistically verify reversed association. Second, empirical experiments and computer simulations are presented to verify the reliability of the reversed association test. Combined together, this study reveals that there is a non-monotonic relationship (i.e., reversed association) in performance for categorization and similarity judgment tasks.

Keywords: Similarity, Categorization.

Introduction

This article examines the cognitive processes underlying categorization and similarity judgments. Although similarity has been known to play a central role in category formation, the role of similarity in categorization has been questioned lately (Hahn & Ramscar, 2001; Hampton et al., 2007). On the empirical side, a number of studies have demonstrated functional independence between similarity and categorization judgments. Some feature information that affects a categorization task is nonetheless ineffective in a similarity judgment task, and vice versa (see Han & Ramscar, 2001). On the basis of this discrepancy, several theorists have proposed that categorization relies on multiple processes – a similarity-based associative process and a rule-based abstract process (Sloman, 1996).

However, this view has been severely criticized on two fronts. First, the empirical evidence for the multiple-processes view of categorization also turned out to be consistent with a similarity-based single process view (Nosofsky & Johansen, 2000; Pothos, 2005). Second, the evidence supporting the multiple-processes view is primarily based on the notion of functional independence, 1 which by itself is not very informative in linking cognitive processes and task performance (Van Orden, 2001). Thus, it is unclear whether similarity and categorization tasks are mediated by the same or different process(es).

In this article, I apply the reversed association theory developed by Dunn and Kirsner (1988) and investigate the relationship between categorization and similarity. In what follows, I will first illustrate the problems with interpreting functional independence as evidence for a multiple-processes view, and then introduce Dunn and Kirsner's reversed association theory as a tool to assess the relationship between task performance and cognitive processes. A statistical method for evaluating reversed association will be presented, and three empirical studies and one computer simulation investigating the dissociation between categorization and similarity will be described. As readers will see, our study implicates a strong possibility that similarity and categorization judgments rely on some unshared cognitive processes, and suggests that the reversed association test is an important and practical method to examine the dissociation of the processes underlying two related tasks.

Dissociation in Similarity and Categorization Judgment

The key evidence for Sloman's two systems of reasoning comes from the behavioral discrepancy underlying categorization and similarity judgments (Sloman, 1996). Rips (1989) provided some of the first empirical evidence for this dissociation. His main argument is that some independent variables affect categorization but not similarity judgments, and vice versa (i.e., functional independence). For example, modifying an internal feature did not change the perceived similarity between original and transformed animals, whereas modifying a surface feature did change the perceived similarity between original and transformed animals. In contrast, modifying an internal feature changed the category membership of original and transformed animals, whereas modifying a surface feature did not change the category membership of original and transformed animals (Rips, 1989). By manipulating central features (Kroskaand & Goldstone, 1996), necessary features (Thibaut, Dupont, & Anselme, 2002), causal features (Ahn & Dennis, 2001), and the frequency of features (Rips & Collins, 1993), other studies demonstrated analogous dissociations.

All of these studies were designed to examine functional independence between categorization and similarity judgments. In the categorization task, participants judged if a target item belonged to one of two designated categories. In the similarity judgment task, participants judged if a target item was similar to one of two designated categories. Participants' responses were analyzed in a 2-alternative forced-choice setting. Two option categories were pitted against each other, so that the selection of one category reflected an operation of one cognitive strategy (e.g., attending primarily to causal features) over the other (e.g., attending equally to causal and non-causal features).

The problem with testing "functional independence" is that its interpretation is not always straightforward. For example, hypothetical data like those shown in Figure 1 may indicate functional independence between two tasks. However, these response patterns can emerge from the two tasks that are different only in their decision thresholds.

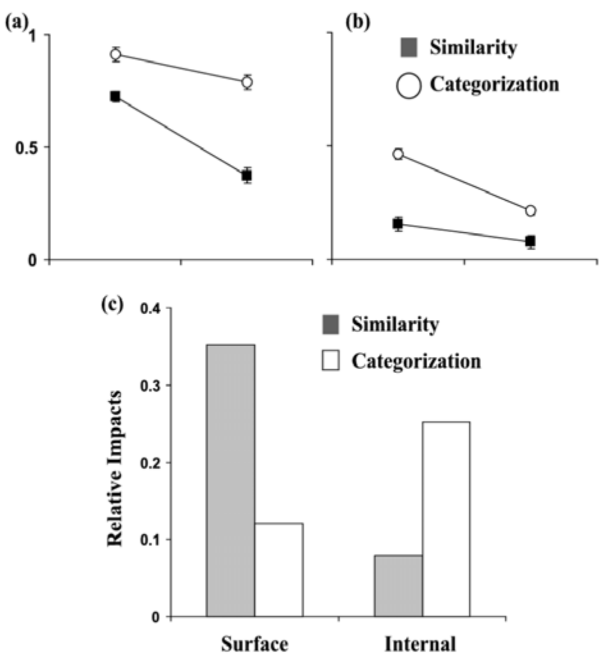

Figure 1: A hypothetical example of *functional independence*. In (a), the performance for a similarity task, but not a categorization task, is affected by a manipulation of surface features. In (b), this relationship is reversed. By measuring the *relative impacts* of the two types of features, a crossed interaction effect can be obtained – (c). For example, in (c) the impact of a superficial feature is larger in a similarity task than in a categorization task. In contrast, the impact of an internal feature is larger in a categorization task than in a similarity task.

Consider the following logistic regression functions (Figures 2a & 2b):

$$p\_task_1 = \frac{1}{(1+\exp(-4+0.1x))} + error \quad -(1)$$

$$p\_task_2 = \frac{1}{(1+\exp(-5.8+0.1x))} + error \quad -(2)$$

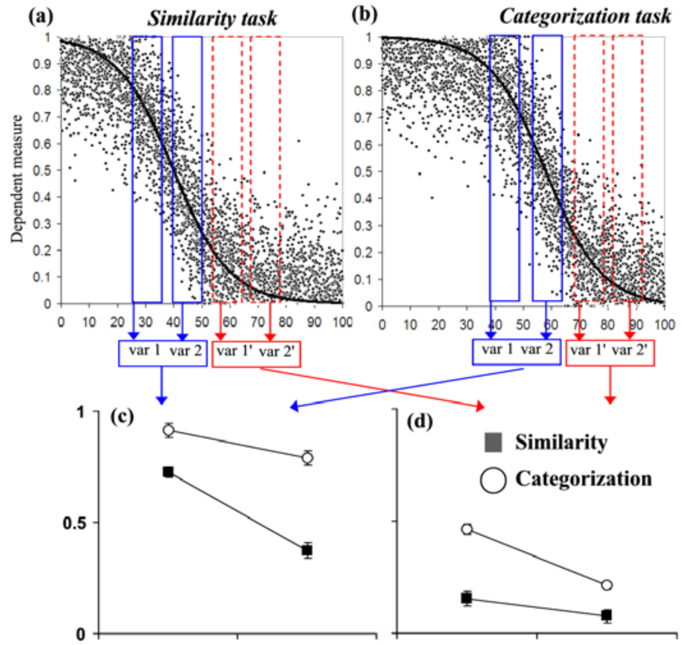

Figure 2: Two logistic regression functions simulating two 2AFC tasks (similarity vs. categorization tasks). The error terms of the two logistic functions are distributed normally with mean 0 and standard deviation 0.15. The 8000 dots shown in (a) and (b) represent hypothetical results from 8000 subjects performing one of the two tasks at a given level of the independent variable (the x-axes). (c) and (d) represent "results" from two hypothetical experiments.

Assume that these functions represent the data collected in two tasks – a similarity judgment task (Figure 2a) and a categorization task (Figure 2b). The x-axes of Figures 2a and 2b represent a hypothetical variable (e.g., the strength of a particular feature dimension) varying from 0 to 100. The y-axes of the figures represent the proportion of selecting one category (e.g., Category A) over the other (Category B) in a similarity task (Figure 2a) or in a categorization task (Figure 2b). The two functions differ only in their decision thresholds. Figure 2b is obtained by shifting Figure 2a by 18 units along the x-axis, and the slopes of these graphs are identical.

Even in this simple setting, a semblance of "functional independence" can emerge from a manipulation of an experimental design. For example, similarity performance is affected by var 1 and var 2 (e.g., modifying surface features), whereas categorization performance is less affected by the same manipulation (Figure 2c). In contrast, the categorization task is affected by var 1' and var 2'(e.g., modifying an internal feature), whereas similarity performance is less affected by the same manipulation (Figure 2d).

This simple observation suggests that a variant of functional independence can emerge even when two tasks differ in their decision thresholds (e.g., one task requires more conservative responses than the other). In other words, demonstrating functional independence is important, but not sufficient in determining whether two tasks stem from the same or different cognitive process(es) (Hampson, 2007; Henson, 2006 for a similar argument).

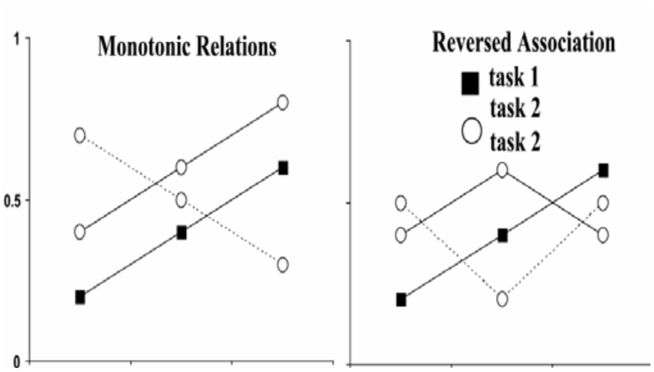

Figure 3: Examples of monotonic relationships between two tasks – (a), and reversed association in task 1 and task 2 – (b).

Dunn-Kirsner's Reversed Association Test

Unlike functional independence, Dunn and Kirsner's (1988) reversed association theory is free from such confusion. The theory is extremely powerful, because it makes only a minimum assumption between a task and a process. In a nutshell, the theory suggests that if the performance for two tasks is not monotonically related, then the two tasks are not mediated by the same process.

For example, in Figure 3, consider three independent manipulations (e.g., manipulating feature information) with respect to two tasks (e.g., similarity and categorization tasks). If the three manipulations change the performance for task 1 monotonically, then the same manipulations should also change the performance for task 2 monotonically (either in an increasing or decreasing fashion – Figure 3a), provided that the two tasks are based on the same process. A violation of this monotonicity is called "reversed association," and it offers strong evidence for the idea that two tasks are not identical in their cognitive processes (Figure 3b) (a summary of a proof is presented in the Appendix A, and see also Dunn & Krisner, 1988 for details).

Despite its theoretical significance, the reversed association theory has not been fully utilized. The problem is its applicability. First, finding reversed association requires at least 6 different conditions (three independent variables applied to two tasks) in a single experiment. It is difficult to implement such a large scale study in many instances (e.g., neuroimaging studies). Second, it is unclear how to verify reversed association statistically. For example, a graph like the one shown in Figure 3b indicates a reversed association, but how can it be verified statistically?

In the next case study, I propose a practical solution for these problems. First, to overcome the need to implement multiple conditions (six conditions), a meta-analytic procedure comparing several independent studies is illustrated. Second, to test the presence of reverse association statistically, a contrast analysis is presented.

Study 1: A reversed association test

The purpose of this case study is to illustrate a statistical procedure to test reversed associations in a series of manipulations made in three independent studies. In one of Rips's (1989) original experiments, participants received a description of an unknown animal called "sorp," and made either a similarity or categorization judgment. Participants in the similarity condition judged whether a "sorp" was *more similar to a bird or an insect*. Participants in the categorization condition received the same stimulus material, but they judged if the same sorp was *more likely to be a bird or an insect*.

Rips's original study consisted of a 2 (categorization vs. similarity judgment tasks) x 2 (accidental vs. essential transformation) factorial design (Rips, 1989, pp.38-42). This design does not allow us to assess the non-monotonicity of the two tasks. In Study 1, we added one more condition to his study, and investigated reversed association in a 2 (categorization vs. similarity judgment tasks) x 3 (accidental vs. essential vs. no-cause conditions) between-subjects design.

The stimuli were 3 different descriptions of an unknown animal called "sorp" whose attributes were later modified in three different ways: accidental, essential, or no-cause. In the Rips study (1989), "sorp" originally had attributes consistent with "birds" (e.g., "has two wings" and "lives in a nest high in the branches of a tree"), but later developed to display characteristics of insects (e.g., "grows two more pairs of legs and clinging upside down to the undersides of tree leaves"). The transformation was caused by hazardous chemicals – the accidental condition – or by a some genetic process – the essential condition.

The accidental and essential conditions in this study were taken directly from the original Rips study. In the other condition, the no-cause condition, the transformation of the animal was described but the actual cause of the transformation was unspecified.

Reversed association test

This section illustrates a meta-analytic procedure to investigate whether the three manipulations implemented separately in the three sets of the study would affect similarity and categorization performance. To examine the impact of the three variables, we first calculated the proportion of selecting one "designated" category in each study. The designated category in this case study was participants' selecting "birds" over "insects." We then calculated the extent to which these proportions exceeded a chance level performance of 0.5 in Z-scores. These Z-scores were converted to effect size *r* and *Fisher Z* to compare the three independent studies. A contrast analysis was applied to assess the overall trend of the influence of the three manipulations. The specific procedure applied in this analysis is as follows:

(a) For the three studies, the proportions of selecting one designated category over the other were calculated. The extent to which the observed proportion *p* exceeds a chance level performance of P_0=0.5 was translated into *Z*-scores with the following formula (p. 13 Fleiss, 1981):

$$Z_i = \frac{(p_i - P_0) - 1/(2n_i)}{\sqrt{\frac{P_0 Q_0}{n_i}}} \quad -(3)$$

where $Q_0 = 1 - P_0$ =0.5, n_i is the sample size of the *i*-th study, and p_i is the proportion observed in a given study.

(b) To compare the three independent studies, these Z-scores were translated into effect size *r*, and then *Fisher Z* (p. 19, Rosenthal, 1984):

$$r_i = \frac{Z_i}{\sqrt{n_i}} \quad -(4)$$

$$Z_{ri} = \frac{1}{2}\log_e\left(\frac{1+r_i}{1-r_i}\right) \quad -(5)$$

where r_i, Z_i, and n_i each represents the effect size, Z-score and sample size of the *i*-th study. Z_{ri} is a transformation of *r* obtained in the *i*-th study (Fisher Z; p. 21 Rosenthal, 1984). This transformation is used to prevent the bias stemming from the distribution of *r* as *r* becomes large.

(c) To assess *reversed association*, two orthogonal contrasts (linear contrast weights=(-1, 0, 1); quadratic contrast weights=(1, -2, 1)) were applied to the Fisher Z's (Z_{ri}), which were obtained separately in the similarity and categorization judgment tasks of the three studies:

$$\frac{\sum \lambda_i Z_{ri}}{\sqrt{\sum \frac{\lambda_i^2}{n_i - 3}}} \quad -(6)$$

λ_i, n_i, and Z_{ri} each represent the contrast weight, sample size, and Fisher Z of the *i*-th study and (6) is known to have the standard normal distribution (p. 80, Rosenthal, 1984; see also Rosenthal & Rosnow, 1985). A statistically significant high score of (6) would support the presence of a particular trend (either a linear or quadratic trend). The presence of *reversed association* can be detected by a significant linear trend in one task and a significant quadratic trend in the other task.

Method

Participants A total of 185 undergraduate students at Texas A&M University participated in this experiment for course credit. These participants were assigned to one of 6 conditions.

Materials & Procedures The stimuli were presented on a piece of paper, on which one of the three different descriptions of "sorp" was shown (accidental, essential, or no-cause descriptions). Participants first read the description, and then answered either a categorization or similarity question. The two questions were identical except for a few words.

(Categorization) Question

Is this sorp more likely to be a bird or an insect? Circle either one below.

Bird Insect

(Similarity) Question

Is this sorp more similar to a bird or an insect? Circle either one below.

Bird Insect

In the three conditions, the unknown animal initially had features consistent with birds and then came to possess the features consistent with insects. The transformation of the bird- to insect features was triggered by an accident (exposure to hazardous chemical waste) in the accidental condition and by a genetic process in the essential condition (Rips, 1989). In the no-cause condition, no specific cause of the transformation was described.

Design The design of this experiment was a 2 (questions: categorization vs. similarity) x 3 (description: accidental, essential vs. no-cause) factorial. The two factors were between-subjects factors.

Results & Discussion

Figure 4 is a graphical representation of the effect sizes generated by the three types of descriptions (accidental, essential, and no-cause). As Figure 4 reveals, a *reversed association* effect is apparent. An application of a linear contrast indicates a significant linear trend in the similarity judgment task; $z = 3.40$, $p < 0.001$, but not in the categorization judgment task; $z = 0.87$, $p = 0.38$. In contrast, a significant quadratic trend was evident in the categorization judgment task; $z = 3.55$, $p < 0.001$ but not in the similarity judgment task; z=0.56, $p > 0.5$.

These results demonstrate a significant linear trend in the similarity task and a significant quadratic trend in the categorization task. However, care should be taken to interpret these results because we do not know the reliability of this statistical procedure. Specifically, if the test still yields a significant reversed association (a significant linear trend in one task and a significant quadratic trend in the other task) even when the two tasks are separated merely by response thresholds (see Figures 2a and 2b), the test would be unreliable. In this regard, it is important to scrutinize the false alarm rate of this test procedure.

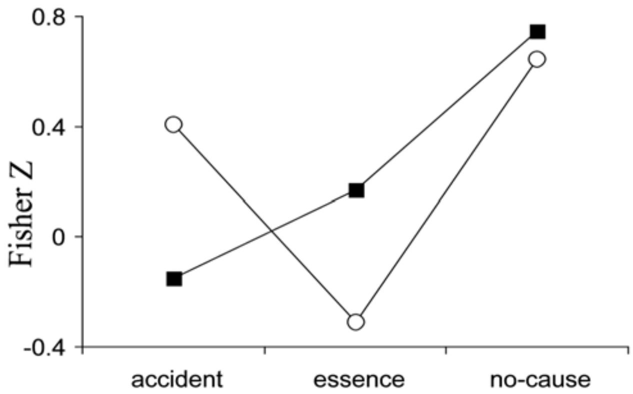

Figure 4: A summary of the results from Study 1.

Study 2: Estimating the false alarm rate of the reversed association test

In Study 2, we conducted 30,000 simulated experiments and examined the false alarm rate of the reversed association test. Specifically, we estimated the probability that statistically significant reversed associations would occur when two tasks were identical except for their response thresholds (i.e., a false alarm rate: False alarm rate = P(A | B), where A = statistically significant reversed associations: B = two tasks differ only in their response thresholds).

The basic procedure of this simulation study was as follows: (1) selecting one logistic regression function from a pool of three, and then modifying it with two different response threshold parameters (Figure 5a and 5b); (2) introducing three intervals (the intervals represent three experimental conditions, such as the accidental, essential, and no-cause conditions in Study 1) to the two functions (Figure 5c); (3) randomly selecting 30 data points from each interval and applying the reversed associated test to the data; (4) repeating this procedure 30,000 times with different parameters, and measuring how often statistically significant reversed association would occur in these 30,000 simulated experiments. If the false alarm rate of the reversed association test is high, then the test is not be very useful.

Method

This simulation study assumes that people's binary responses, such as selecting one category from two options (e.g, birds vs. insects), can be modeled by a logistic regression function. A logistic regression function is defined by two parameters, its slope and inflection point.

$$y = \frac{1}{(1+\exp(-b+ax))} + error \quad \text{-- (7)}$$

Parameter a in (7) corresponds to the slope of the function (see Figure 5a), and $-b/a$ corresponds to the point at which the probability of selecting one category is equal to the other (P(cate_A)=P(cate_B)=0.5). I call this point "the inflection point" of a function.

Figure 5b shows two logistic regression functions with two different inflection points.

$$p\_task_1 = \frac{1}{(1+\exp(-4+0.1x))} + error \quad \text{-- (8)}$$

$$p\_task_2 = \frac{1}{(1+\exp(-6+0.1x))} + error \quad \text{-- (9)}$$

In (8), chance level performance (P(cate_A)=P(cate_B)=0.5) occurs at a level of 40 ($-b/a$=40), whereas in (9), chance level performance occurs at a level of 60 ($-b/a$=60). To simulate different response thresholds applied in similarity and categorization tasks, we generated 200 random pairs of inflection points ($-b/a$) with a restriction that the parameter $-b/a$ ranged from 0 to 100. To simulate responses made by individual participants, error terms were added to the two functions and 4000 random data points were generated for each of the two functions (Figure 5c).

$$p\_task_1 = \frac{1}{(1+\exp(-b+ax))} + error \quad \text{-- (10)}$$

$$p\_task_2 = \frac{1}{(1+\exp(-b'+ax))} + error \quad \text{-- (11)}$$

The error terms were assumed to be normally distributed with mean 0 and standard deviation 0.15. To simulate the three variables implemented in each experiment (e.g., accidental, essential, and no-cause conditions in Study 1), we dividend the x-axis into 10 intervals ranging from 0 to 100 (see the x-axis of Figure 5c), and randomly selected 50 triads of intervals with the following restrictions: 0<interval_1< interval_2< interval_3<100.

Thirty samples were selected randomly from each interval, from which the effect sizes were calculated, and contrast analyses employing linear and quadratic weights were implemented. This procedure was applied to three different logistic regression functions with three different slopes (a=0.1, 0.06, 0.04, and see Figure 5). A "statistically signification reverse association effect" was measured by the presence of a significant linear contrast in one task and a significant quadratic contrast in the other task. To assure that our statistics did not miss indications of false alarms, we set the alpha level to be particularly lenient; alpha = 0.1 (one tailed).

In summary, this simulation study employed the following steps. Step 1: we selected one logistic regression function from a pool of three and generated 200 pairs of functions that had different response threshold parameters ($-b/a$); Step 2: 50 different randomly selected triads of variable intervals were implemented to every pair of the functions and 30 random samples were taken from each interval; Step 3: contrast analyses were applied to every pair of the functions (Figure 5), and the presence of a statistically significant reversed association effect was tallied; Step 4: Steps 1-3 were repeated for the other two remaining logistic regression functions.

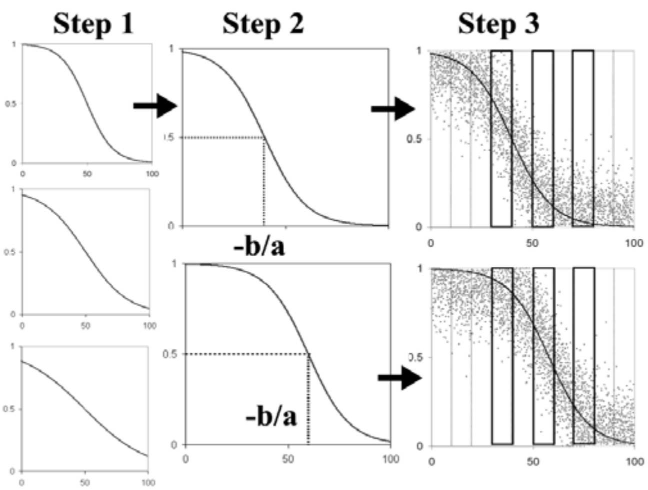

Figure 5: A graphical representation of the procedure used in the simulation study (Study 2).

Results

The estimated false alarm rate of the reversed association test was quite small. We found only one occasion (out of 30,000 simulated experiments) in which the test resulted in a significant reversed association. To assure that the low false alarm rate did not come from the arbitrary selection of the parameters, we conducted additional studies. First, we raised the alpha level from 0.10 to 0.15. Second, the standard deviation of the error term was also changed from 0.15 to 0.20. Even with these modifications, the false alarm rate remained the same.

Discussion

The results from the two studies have shown that the *reversed association* test could be a promising tool in detecting the dissociation between two tasks. But in order to use this tool effectively, it is crucial to know its limitations. This section summarizes important limitations of the reversed association test.

(a) This is NOT a necessary and sufficient test. The purpose of assessing linear and quadratic trends in two tasks is to statistically verify monotonic and non-monotonic relationships in two tasks. Assessing a liner trend is one way of capturing the monotonicity of a function; however, there are many nonlinear monotonic functions (e.g., $y = x^3$). This means that when the test does not show significant effects, it can NOT be concluded that there is NO reversed association between two tasks. In this sense, it is important to show first whether the data conform to the general pattern of *reversed association* (e.g., by means of plotting) and then apply the statistical test.

(b) The two tasks should be compared in a tightly controlled condition. Subtle differences in the procedures, participants, instructions and stimuli are likely to result in significant differences in the performance for the categorization and similarity judgment tasks. In this regard, the two tasks should be compared under a tightly controlled uniform condition whenever possible.

(c) The test does not identify the underlying "processes" that separate two tasks. The reversed association test is useful to detect the "dissociation," but the test does not tell what "process" contributes to the dissociation. In other words, the test cannot identify the cause of the "dissociation."

References

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review, 95*, 91-101.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.

Hahn, U., & Ramscar, M. (Eds.) (2001). *Similarity and categorization*. Oxford, UK: Oxford University Press.

Hampton, J. A. (2007). Typicality, graded membership, and vagueness. *Cognitive Science, 31*, 355-384.

Hampton, J. A., Estes, Z., & Simmons, S. (2007). Metamorphosis: Essence, appearance, and behavior in the categorization of natural kinds. *Memory & Cognition, 35*(7), 1785-1800.

Henson, R. (2006). Forward inference using functional neuroimaging: dissociations versus associations. *Trends in Cognitive Sciences, 10*, 64-69.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review, 7*, 375-402.

Pothos, E. M. (2005). The rules versus similarity distinction. *Behavioral and Brain Sciences, 28*, 1-49.

Rips L. J. (1989) Similarity, Typicality, and Categorization, In S. Vosniadou andN. Ortony (Eds.) (pp. 21-59), *Similarity and Analogical Reasoning*. Cambridge: Cambridge University Press.

Rosenthal, R. (1984). *Meta-analytic procedures for social research*. London: Sage Publications.

Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis in behavioral research: Focused comparisons in the analysis of variance*. New York: Cambridge University Press.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science, 25*, 111-172.