

Memory & Cognition
2006, 34 (7), 1452-1463

Testing the semantic differential as a model of task processes with the implicit association test

MAGGIE J. XIONG, GORDON D. LOGAN, and JEFFERY J. FRANKS
Vanderbilt University, Nashville, Tennessee

In this study, we examined the hypothesis that semantic judgment tasks share overlapping processes if they require processing on common dimensions but not if they require processing on orthogonal dimensions in semantic space (Osgood, Suci, & Tannenbaum, 1957). We tested the hypothesis with the implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998) in three experiments. Consistent with the hypothesis, IAT effects (costs in reaction time because of incompatible response mapping between associated judgment tasks) occurred consistently when judgment tasks tapped into common semantic dimensions, whereas no IAT effect appeared when judgment tasks entailed processing on orthogonal semantic dimensions.

Similarity is a basic theoretical construct in accounts of cognition and memory. The facilitation or inhibition in the processing of a stimulus has been described as a continuous function of stimulus and response similarities (Osgood, 1949). The effects of task similarity are apparent in myriad cognitive phenomena, such as costs in task switching (Arrington, Altmann, & Carr, 2003), trade-offs between tasks in the dual-task paradigm (Navon & Gopher, 1979), and transfer among explicit and implicit memory tasks (Roediger, Weldon, & Challis, 1989).

For task similarity to be an empirically useful construct, one must be able to specify precisely why, or in what respects, things are similar (Medin, Goldstone, & Gentner, 1993), and this specification can be difficult because perceived similarity between situations varies with the context and the task (Murphy & Medin, 1985; Tversky, 1977). When the tasks under investigation are oriented to perceptual processing, the similarity between a pair of tasks can be relatively easily defined in terms of the perceptual dimensions accessed, such as color and brightness, or in terms of presentation modalities, such as visual and auditory. When the tasks of interest draw heavily on conceptual processing, defining task similarity becomes a more difficult problem.

Approaches to conceptual similarity between tasks include the domain-specific attributes approach (Thompson-Schill & Gabrieli, 1999; Vriezen, Moscovitch, & Bellos, 1995) and the competitive/noncompetitive access theory of conceptual knowledge (Vaidya et al., 1997). The domain-specific attributes approach suggests that semantic tasks are similar if they access the same semantic domain

(e.g., structural vs. functional domain). The competitive/noncompetitive access theory distinguishes between conceptual tasks that arouse multiple response candidates and conceptual tasks that elicit a unique response. While these accounts are reasonable descriptions of the results reported by these authors, they fail to account for other patterns of data, as presented in our recent reports of repetition priming, that seem to reflect different aspects of conceptual similarity (Franks, Bilbrey, Lien, & McNamara, 2000; Xiong, Franks, & Logan, 2003). In particular, Xiong et al. presented a case for the importance of semantic differential representations of conceptual similarity.

In the present study, we further evaluated the semantic differential (Osgood, Suci, & Tannenbaum, 1957) as a functional model for defining similarity relations between semantic judgment processes in the semantic judgment task space. Task similarity in this case is defined with respect to the dimensions in semantic space that are accessed by the semantic judgment tasks. In the present work, we used the implicit association test (IAT) paradigm (Greenwald, McGhee, & Schwartz, 1998) to assess hypotheses generated by the semantic differential account of conceptual similarity among tasks.

The semantic differential (Osgood et al., 1957) is a framework for measuring the connotative meaning of words in semantic space. Semantic space is defined by a finite number of orthogonal semantic dimensions along which the connotative meaning of words may vary. Words are represented as points in space, and the direction and magnitude of a word's coordinate on each dimension corresponds to the nature (positive or negative on that dimension) and intensity of the reaction elicited by the word. To identify major semantic dimensions, Osgood et al. had subjects rate the meanings of words on a set of "semantic scales," such as *good/bad* and *large/small*. Osgood et al. then performed factor analysis on correlations between the semantic scales. The first three semantic factors (dimensions) extracted were the "evaluative," "potency," and

This research was supported by National Science Foundation Grant BCS 0133202. Correspondence concerning this article should be addressed to M. J. Xiong, Department of Psychology, Vanderbilt University, Nashville, TN 37203 (e-mail: maggie.xiong@vanderbilt.edu).

Note—This article was accepted by the previous editorial team, when Colin M. MacLeod was Editor.

"activity" factors, in order of the proportion of variance accounted for (see also Bentler & LaVoie, 1972; Tzeng, 1975). These results were replicated qualitatively in our lab with a new set of words and a subset of Osgood et al.'s semantic scales. Factor loadings of the semantic scales mentioned in the present study, when available, are listed in Table 1, including both loadings from Osgood et al.'s Analysis I and the replication from our lab.

Semantic judgments, sometimes referred to as *semantic classifications*, involve making judgments about the meaning of words—for example, making a *good/bad* judgment about *rose*. Xiong et al. (2003) suggested that the similarity between semantic judgments on words could be determined by whether these judgments require processing on the same semantic dimensions, which is in turn decided by whether the corresponding semantic scales are loaded on the same semantic factor(s). Semantic judgments should be similar if the corresponding semantic scales have significant loadings on the same semantic factor(s). For instance, both *pleasant/unpleasant* and *valuable/worthless* are loaded on the evaluative factor, so *pleasant/unpleasant* and *valuable/worthless* judgments are considered similar to each other. Semantic judgments should be dissimilar if the corresponding semantic scales have significant loadings on different semantic factor(s). As an example, *pleasant/unpleasant* is loaded primarily on the evaluative factor, whereas *strong/weak* is loaded primarily on the potency factor, so there should be minimal overlapping processes between *pleasant/unpleasant* and *strong/weak* judgments.

Xiong et al. (2003) employed a repetition priming paradigm (for reviews, see Roediger & McDermott, 1993, and Schacter, 1987) to test the semantic differential account of similarity between semantic judgment processes. The subjects performed a semantic judgment task on words during the acquisition. At test, they performed the same semantic judgment or a different one on the old set and on a new set of words. Repetition priming was the reduction in reaction

time (RT) on the old set relative to the new set of words at test. Consistent with the hypotheses, Xiong et al. detected various amounts of repetition priming when the acquisition and test semantic scales were loaded on the same semantic factor(s), and they detected no reliable repetition priming when the acquisition and test semantic scales were loaded primarily on orthogonal semantic factors.

However, further work in our lab revealed significant repetition priming between semantic judgments that were supposedly orthogonal to each other and therefore should involve little or no overlapping processes. *Good/bad* and *large/small* were loaded primarily on the evaluative and potency factors, respectively (Osgood et al., 1957). After they had performed *good/bad* judgment on a set of words, subjects were significantly faster at making *large/small* judgment on this set of words than on a new set of words. Similarly, when subjects had made *large/small* judgment on words, they were faster at making *good/bad* judgment on these words than on new words.

In order to accommodate this observation, the semantic differential account of task similarity would have to be modified or augmented. One possible augmentation would be to assume that repetition priming could be influenced by factors other than task similarity. Similarity between the required acquisition and test tasks may lead to facilitation at the test phase, yet facilitation at test does not necessarily indicate similarity between the required acquisition and test tasks. Franks et al. (2000) speculated that characteristics of the words or the required judgment or both might cue subjects into making certain judgments on words even if the experimental instruction did not ask for such judgments. In the case of *good/bad* and *large/small* judgments, when presented with the word *ocean*, subjects might automatically encode it as being *large*, even though the required acquisition task was *good/bad* judgment. When subjects were later asked to make *large/small* judgment at test, repetition priming for the item *ocean* would be observed. This type of repetition priming should not be

Table 1
Semantic Scales' Loadings on the Semantic Factors

Semantic Scale	Semantic Factors (Osgood et al., 1957)					Semantic Factors (Replication)				
	I	II	III	IV	h^2	I	II	III	IV	h^2
<i>active/passive</i>	.14	.04	.59	-.02	.37	-.02	.17	.73	.10	.58
<i>fast/slow</i>	.01	.00	.70	-.12	.50	-.03	.09	.68	-.01	.47
<i>good/bad</i>	.88	.05	.09	.09	.79	.94	.04	-.01	.02	.92
<i>hard/soft</i>	-.48	.55	.16	.21	.60	-.41	.21	.07	-.72	.77
<i>large/small</i>	.06	.62	.34	.04	.51	.03	.84	.12	-.07	.87
<i>pleasant/unpleasant</i>	.82	.05	.28	.12	.77	.94	.03	-.08	.14	.92
<i>strong/weak</i>	.19	.62	.20	.03	.46	.09	.60	.45	-.28	.65
<i>valuable/worthless</i>	.79	.04	.13	.00	.64	.80	.26	.06	-.10	.74
% of total variance	34	8	6	2	49	31	12	11	6	66
% of common variance	69	15	13	3	100	48	18	17	9	100

Note—For each semantic scale, the table lists factor loadings from Osgood et al.'s (1957) Analysis I on the left and factor loadings from our lab's replication on the right. The first three factors accounted for most of the explained variance and were the most consistent across similar studies (Bentler & LaVoie, 1972; Tzeng, 1975). They corresponded respectively to the evaluative, potency, and activity dimensions of semantic space. Unlike Analysis I, our replication yielded six instead of four factors. Commonalities (h^2) and total percent of explained variance from the replication are based on all six factors.

taken as an indicator of similarity between *good/bad* and *large/small* judgment processes.

Whether other factors besides task similarity might contribute to repetition priming is a topic that warrants separate study. Instead of hunting down these other factors, in the present study, we described an alternative method for evaluating the semantic differential account of similarity between semantic judgments.

The IAT is a procedure for assessing the strength of association between concepts (Greenwald et al., 1998; for reviews, see De Houwer, 2002, and Greenwald & Nosek, 2001). Originally developed to detect implicit attitudes, the IAT procedure investigated associations between "target concepts," such as *flower* and *insect*, and evaluative "attribute concepts," such as *pleasant* and *unpleasant*. Two categorization tasks—*flower/insect* and *pleasant/unpleasant* categorizations—were randomly mixed in the same block (see Figure 1A). When subjects were presented with items from all categories successively (e.g., *aster*, *ant*, *caress*, *abuse*), they responded faster when positively associated categories (i.e., *flower* and *pleasant*, *insect* and *unpleasant*) shared response keys, as compared with when negatively associated categories (i.e., *insect* and *pleasant*, *flower* and *unpleasant*) shared response keys. These two ways of mapping categories to responses were referred to as the *compatible* and *incompatible* mappings, respectively. The *IAT effect* was the difference in average RT between the compatible and incompatible response mappings. Karpinski and Hilton (2001) argued that IAT effects reflect the environmental association between target concepts and evaluative attributes. They

manipulated the frequency of co-occurrence of target concept stimuli and evaluative attribute stimuli before administering a standard *flower/insect* IAT. IAT effects were either strengthened or weakened depending on whether these two types of stimuli co-occurred in a stereotypical way or the opposite.

The semantic differential (Osgood et al., 1957) was originally developed on the basis of just such co-occurrences of representational processes. The fact that IAT effects are sensitive to the frequency of co-occurrences suggests that IAT should be an appropriate tool for assessing semantic differential relations in semantic judgments. If target concepts and attribute concepts are replaced with semantic scales, IAT effects should reflect associations between semantic judgment processes. In addition to providing information about the strength of association between semantic scales, IAT allows us to assess the direction of association between semantic scales. In the semantic differential, scales point to positive and negative ends of semantic dimension(s). IAT effects should reveal whether the ends of these semantic scales are aligned or misaligned. Repetition priming experiments (Xiong et al., 2003) have not tested this aspect of the semantic differential framework.

If semantic judgment tasks are associated when the corresponding semantic scales have loadings on common semantic factor(s), IAT effects should be observed when these judgments are randomly mixed in the same blocks. Responses should be made faster when positively associated ends of the semantic scales are mapped to the same keys (compatible mapping) than when negatively associated ends of the semantic scales are mapped to the same keys (incompatible mapping). For instance, *large/small* and *hard/soft* are both positively loaded on the potency factor. Responses should be made faster when *large* and *hard* responses and *small* and *soft* responses are mapped to the same keys than when *large* and *soft* responses and *small* and *hard* responses are mapped to the same keys.

On the other hand, when semantic scales are loaded on orthogonal semantic factors, there should be no association between the corresponding judgments, and no IAT effect should be observed. *Good/bad* and *large/small* are loaded on the evaluative and potency factors, respectively. When *good/bad* and *large/small* judgments are randomly mixed in the same block, it should make no difference whether *good* and *large* responses and *bad* and *small* responses or *good* and *small* responses and *bad* and *large* responses are mapped to the same keys.

EXPERIMENT 1

In Experiment 1, we examined IAT effects between three pairs of semantic judgments: *good/bad* and *hard/soft*, *large/small* and *hard/soft*, and *good/bad* and *large/small*. Judgments in the first two pairs are associated, while judgments in the third pair are orthogonal to each other according to the semantic differential approach. *Good/bad* is positively loaded on the evaluative factor. *Hard/soft* is negatively loaded on the same factor. *Good* and *soft* responses and *bad* and *hard* responses should be

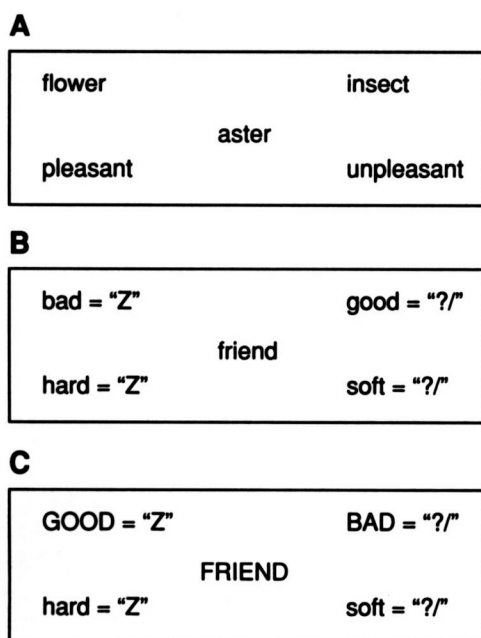


Figure 1. IATs. (A) Example of a typical *flower/insect* IAT. (B) Example from Experiment 1. (C) Example from Experiment 2.

compatible with each other. In contrast, *good* and *hard* responses and *bad* and *soft* responses should be incompatible. Both *hard/soft* and *large/small* are positively loaded on the potency factor. *Hard* and *large* responses and *soft* and *small* responses should be compatible, whereas *hard* and *small* responses and *soft* and *large* responses should be incompatible with each other. *Good/bad* and *large/small* are loaded on the orthogonal evaluative and potency factors, respectively. RTs should be similar whether *good* and *large* responses and *bad* and *small* responses or *good* and *small* responses and *bad* and *large* responses were mapped to the same keys. See the IAT phase in Figure 2 for an illustration of the compatibility between response mappings for these pairs of semantic judgment tasks.

In a typical *flower/insect* IAT, the stimulus item can help determine which categorization task to perform in a trial: If it is a type of *flower* or *insect*, do a *flower/insect* categorization; otherwise, do the *pleasant/unpleasant* categorization. In an IAT with semantic judgments, a stimulus word by itself cannot necessarily specify the required judgment, because either judgment would be legitimate on the word. To get by this problem, establish strong word-judgment relations. The subjects in Experiment 1 were trained to associate semantic judgments with a small set of stimulus words before they proceeded to the IAT.

The subjects then decided which semantic judgment to perform on a word by the trained association between the word and the judgment. Figure 2 shows the experimental block sequences in the training and IAT phases.

Method

Subjects. Altogether, 25 subjects participated in this experiment. They were either students at Vanderbilt University who participated for course requirement or student volunteers who were paid \$5 for participation. All subjects were native speakers of English. They had normal or corrected-to-normal vision. All subjects signed a consent form to participate in the study. After we had gathered data from 24 subjects, we tested an additional subject to replace the data from a subject who had extremely long RTs (more than 2.5 SDs away from the mean RT for all subjects).

Stimuli. Twenty-four nouns were used, 8 for each judgment. They are listed in Appendix A for reference. The items were selected from our replication of the semantic differential. In the replication, subjects rated words on 7-point (1–7) semantic scales. These ratings were subsequently standardized for each semantic scale. If we consider *good*, *hard*, and *large* as the “positive” ends of the scales and *bad*, *soft*, and *small* as the “negative” ends of the scales, half of the selected items were rated clearly positive and half negative on the relevant semantic scales. Across the three semantic scales in this experiment, the average *z*-scores of the positive and negative items were 1.70 and –1.97, respectively. *z*-scores of items ranged from –1.00 to 0.92 on the irrelevant semantic scales. In other words, the items were rated more extreme and should consequently be easier to judge, in terms of their respective judgment than the other two judgments in the experiment. The items had an average word frequency of 105 per million, with an *SD* of 128 per million (Kučera & Francis, 1967). Frequencies did not differ between items for different judgments. The item length varied from four to seven letters.

Procedure. All manipulations in the experiment were within subjects. The subjects were tested individually. Task instructions and materials were presented on 8088 personal computers in 80-column lowercase font, in the default display color of green or orange. Each letter was about 2.5 × 2.5 mm or 2.5 × 5 mm (with extender) in size. Instructions included examples and warned the subjects of possible response mapping changes. The subjects initiated the blocks of trials. Items were presented in the middle of the computer screen for judgment one at a time. The subjects indicated different responses by pressing the “Z” or the “?” key on the keyboard. For both training and IAT phases, a reminder of which key corresponded to which judgment remained on the screen throughout each block, flanking the stimulus items (see Figure 1B).

During the training phase, the subjects performed three blocks of semantic judgments, one for each type of judgment. The order of the blocks was randomized for each subject. Eight words were presented four times in each block. The presentation order of items was pseudorandomized. There was at least 1 item and were at most 15 items between repetitions of words, producing an average repetition lag of 7 items. *Good*, *hard*, and *large* responses were always mapped to a subject’s nondominant hand in the training phase.

The subjects then moved on to the IAT phase, where pairs of semantic judgments were mixed in the same block. Figure 1B illustrates what a trial looked like in this phase. The subjects were instructed to make a response according to the trained judgment for that item. Before every experimental block, the subjects were given the specific response mapping and a practice block. All 16 words for both semantic judgments were presented once in the practice block. There were two experimental blocks for each pair of semantic judgments. The words were presented twice in experimental blocks, resulting in 32 experimental trials in each block. The presentation order of items was pseudorandomized. There was at least one item and were at most 31 other items before a word was repeated. The average repetition lag was 15 items. In one of the experimental blocks,

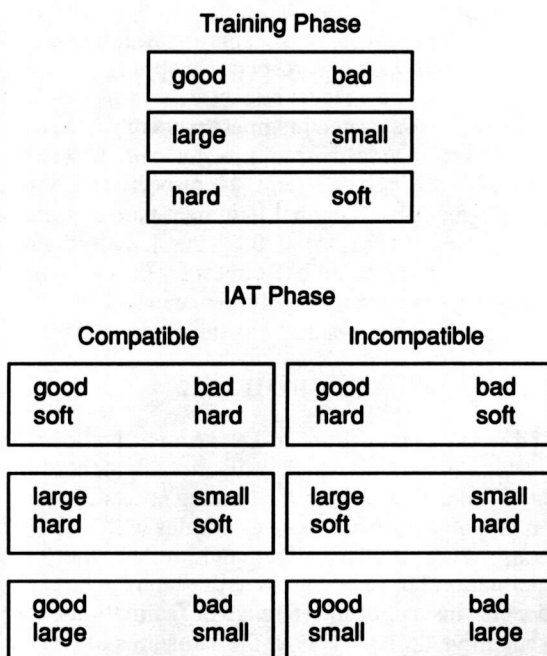


Figure 2. Experiment 1 design. Each box represents an experimental block of semantic judgments. The subjects completed the training phase before proceeding to the IAT phase. During the IAT phase, the subjects finished both the compatible response mapping blocks and the incompatible response mapping blocks for a pair of semantic judgment tasks before moving on to another pair. For the semantic judgments pair of *good/bad* and *large/small*, the labeling of “compatible” versus “incompatible” response mapping in the figure is arbitrary because no reaction time (RT) difference was predicted between these blocks.

the response mapping for both judgments was the same as in the training phase. In the other block, the response mapping for one of the judgments was reversed. This produced the compatible and incompatible response mapping blocks. To be more specific, for *good/bad* and *hard/soft* judgments, the response mapping was incompatible when both judgments had the same mapping as in the training phase. For *large/small* and *hard/soft* judgments, the response mapping was incompatible when one of the judgments had the opposite response mapping as in the training phase.

There were a total of six experimental blocks for the three pairs of judgments in the IAT phase. The subjects completed both the compatible and the incompatible mapping blocks for a pair of semantic judgments before they went on to another pair of judgments. The order of the judgment pairs, as well as the order of compatible and incompatible mapping blocks for each pair, was counterbalanced across subjects. Within a pair of semantic judgments, the judgment that received the reversed response mapping was also counterbalanced.

Each trial began with the presentation of an item. The item remained on the screen until the subject responded. The interval between a subject's response and the presentation of the next item was 500 msec. An "X" appeared for 300 msec below the item if the subject made a wrong response. The presentation order, response accuracy, and RT were recorded for all but practice trials.

Results and Discussion

Mean error rates were 3% in the training phase and 5% in the IAT phase. Detailed error rate information for the IAT phase can be found in Figure 4. Only RTs from correct responses were analyzed. Separate ANOVAs were performed on data from the training phase and that IAT phase. An alpha level of .05 was used for all ANOVAs in the present study. For the training phase, the independent variables were type of semantic judgment and the number of repetitions. Response latencies for the training phase are displayed as a function of these variables in Figure 3. For each judgment, mean RT decreased over successive presentations of the items (Logan, 1990). Both main effects of judgment type and repetition status were significant [$F(2,46) = 5.57$, $\eta_p^2 = .20$, and $F(3,69) = 81.87$, $\eta_p^2 = .78$, respectively]. There was no significant interaction between the two factors.

Data from the IAT phase showed a pattern exactly as predicted by the semantic differential account of similarity between semantic judgments. As illustrated in Figure 4, when semantic scales in the IAT had significant loadings on the same semantic factor(s), RTs were over 200 msec

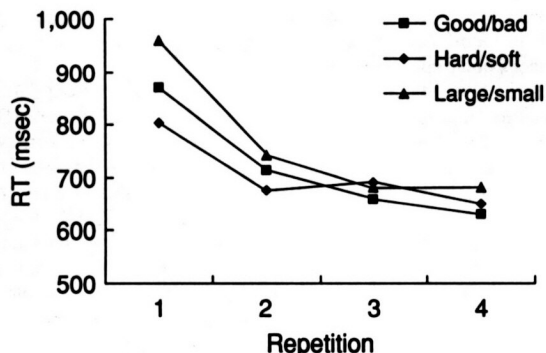


Figure 3. Reaction times (RTs) from the training phase of Experiment 1.

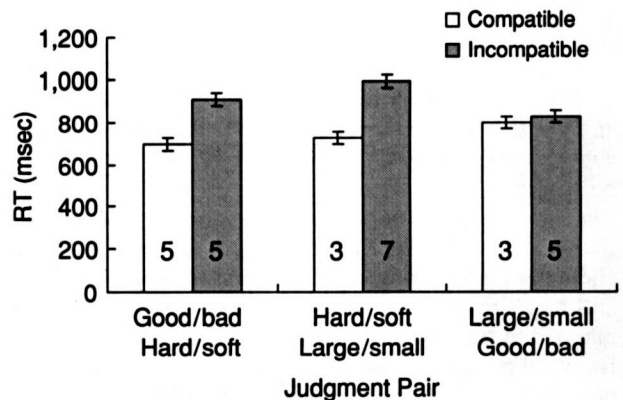


Figure 4. Reaction times (RTs) from the IAT phase in Experiment 1. Error bars stand for standard errors from the ANOVA. Numbers on the bars are error rates.

longer with the hypothesized incompatible response mapping than with the hypothesized compatible response mapping between the judgments. In contrast, when the semantic scales were loaded on orthogonal semantic factors, there was only a 30-msec difference between the two response mapping conditions. A 3 (judgment pair) \times 2 (response mapping) repeated measures ANOVA showed that both main effects were significant [$F(2,46) = 3.47$, $\eta_p^2 = .13$, and $F(1,23) = 70.43$, $\eta_p^2 = .75$, respectively]. More importantly, there was a significant interaction between judgment pair and response mapping [$F(2,46) = 7.92$, $\eta_p^2 = .26$]. The effect of response mapping was highly reliable for the associated judgment pairs of *good/bad* and *hard/soft* and of *hard/soft* and *large/small* [$F(1,46) = 23.61$ and 36.22 , η_p^2 s = .34 and .44, respectively], but was negligible for the orthogonal judgment pair of *good/bad* and *large/small* [$F(1,46) = 0.45$, n.s.]. Paired-sample *t* tests were performed on IAT effects for the two semantic judgments within a pair. No difference in IAT effects was observed between any paired semantic judgments.

EXPERIMENT 2

The training procedure in Experiment 1 allowed the subjects to determine which semantic judgment to make in the IAT phase. However, the training procedure also enabled the subjects to associate a stimulus with a particular response. Thus, the IAT effect could reflect interference or facilitation between cognitive decisions and response tendencies instead of interference or facilitation between two cognitive decisions about the semantics of words. To assess cognitive conflict between two semantic judgment processes, in Experiment 2, we introduced a new variation on the IAT procedure that bypassed the training phase: The required judgment was cued by the letter case in which a stimulus item was presented.

The basic design of Experiment 2 was similar to the IAT phase of Experiment 1. The differences were as follows: (1) Instead of trained associations, different letter cases were used to cue the subjects as to which semantic judgment to perform in a trial. As illustrated in Figure 1C,

the two judgments were assigned to different letter cases in each block. The letter case in which an item was presented specified the judgment to be performed in that trial. (2) All items appeared only once in the experiment. (3) No error feedback was provided.

Method

Subjects. Altogether, 25 student volunteers at Vanderbilt participated in this experiment. They were paid \$5 for participation. All subjects were native speakers of English. They had normal or corrected-to-normal vision. All subjects signed a consent form for participation. After we had gathered data from 24 subjects, we tested an additional subject to replace the data from a subject who had extremely long RTs (more than 2.5 SDs away from the mean RT for all subjects).

Tasks and Stimuli. The same pairs of semantic judgments as in Experiment 1 were used. The stimulus items included 216 nouns, 24 of which were used for filler items. The experimental items are listed in Appendix A. There were 64 experimental items for each judgment task. Most of the items came from our replication of the semantic differential. All items were selected so that they should be easier to judge in terms of their respective judgment task than the other two judgments in the experiment. For every subject, experimental items for each judgment task were randomly divided into four subsets of 16 items each. Because each block in the experiment contained a random mix of two judgments, each block required two subsets of items, one for each judgment task. For every judgment, there were equal numbers of items that could be ascribed to each end of the corresponding semantic scale. All items were presented only once in the experiment. The average word frequency for the experimental items was 89 per million, with an SD of 121 (Kučera & Francis, 1967; *bagel*, *daisy*, *pimple*, *quilt*, and *termite* do not have corresponding entries in the Kučera and Francis norms). Word frequencies did not differ between items of different judgments. The item length ranged from four to seven letters.

Procedure. There were two experimental blocks for every pair of semantic judgments. In one of the blocks, *good*, *hard*, or *large* responses were mapped to a subject's dominant hand. In the other block, response mapping for one of the two judgments in a pair was reversed. This manipulation produced the compatible and incompatible mapping blocks. There were altogether six blocks, with 36 trials in each block, the first 4 trials being filler trials. The presentation order of experimental items was randomized. Experiment 2 followed the same counterbalancing scheme as in Experiment 1.

At the beginning of each block, the subjects were told that one judgment was assigned to the uppercase font and the other to the lowercase font for that block. The response key reminder flanking the stimulus item showed the letter case assignment. Items for each judgment task were presented in the corresponding letter case. The assignment of letter cases to judgments was randomized for every block. The testing procedure was the same as in Experiment 1, except that no error feedback was provided.

Results and Discussion

Because there were no predetermined correct responses for the judgments, error rate was calculated by comparing individual subject's response to an item against the normative response across subjects for that item. Mean error rate was 10%. Only RTs from correct responses were analyzed. Error rates and mean RTs for various conditions are displayed in Figure 5. For associated judgment pairs, responses took about 100 msec longer with incompatible response mapping than with compatible response mapping between the judgments. For the orthogonal judgment pair, there was only a 15-msec difference between the two response mapping conditions. A 3 (judgment pair) \times 2 (response mapping)

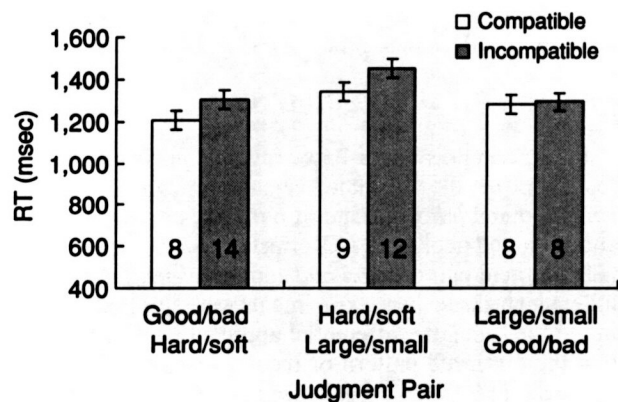


Figure 5. Reaction times (RTs) in Experiment 2. Error bars stand for standard errors from the ANOVA. Numbers on the bars are error rates.

repeated measures ANOVA showed that both main effects of judgment pair and response mapping were significant [$F(2,46) = 9.59$, $\eta_p^2 = .29$, and $F(1,23) = 6.83$, $\eta_p^2 = .23$, respectively]. Although the interaction between judgment pair and response mapping did not reach significance level [$F(2,46) = 1.38$, $p = .26$, $\eta_p^2 = .06$], planned comparisons showed that the effect of response mapping was statistically reliable for *good/bad* and *hard/soft* judgments and for *hard/soft* and *large/small* judgments [$F(1,46) = 4.86$ and 6.32 , $\eta_p^2 = .10$ and $.12$, respectively]. No effect of response mapping was found for the orthogonal judgment pair of *good/bad* and *large/small* [$F(1,46) = 0.12$, n.s.]. The effects of response mapping did not differ between the two judgments within pairs according to paired-sample t tests.

The same pattern of results emerged from an item analysis. Because items for each judgment task were randomly assigned to compatible versus incompatible conditions across subjects, response to an item in the compatible condition could be compared with response to the same item in the incompatible condition. There were on average five to six observations per item in either response mapping condition. As indicated in the "Tasks and Stimuli" section, each item could potentially appear in four out of six blocks in the experiment. We thus performed t tests rather than ANOVAs for the item analysis. For the associated *good/bad* and *hard/soft* judgment pairs and for *hard/soft* and *large/small* judgment pairs, items were responded to significantly slower in the incompatible response mapping condition than in the compatible response mapping condition [$t(127) = 2.07$, and $t(126) = 2.66$, respectively]. For the orthogonal *good/bad* and *large/small* judgments, items were responded to with similar latencies in the two response mapping conditions [$t(126) = 0.50$, n.s.].

Experiment 2 replicated the results of Experiment 1 with a different procedure. Consistent with the semantic differential account of similarity between semantic judgments, IAT effects were detected only when semantic scales representing the judgments were loaded on common semantic factors. By eliminating the training phase and presenting only unique items, in Experiment 2 we demonstrated that

opposition between cognitive processes in semantic judgments was sufficient to produce IAT effects.

EXPERIMENT 3

In Experiments 1 and 2, we found significant IAT effects between the associated semantic judgment pairs of *good/bad* and *hard/soft* and of *hard/soft* and *large/small*, and we found negligible IAT effects between the orthogonal judgment pair of *good/bad* and *large/small*. Although judgment tasks in these experiments were selected according to the semantic differential approach, it was possible that the observed pattern of results was specific to the particular judgments used, instead of general to the whole domain of semantic judgments. To address the generalizability issue, we conducted Experiment 3 with a new set of judgments, *strong/weak*, *active/passive*, and *fast/slow*, using the same procedure as in Experiment 2.

According to the semantic differential approach, *strong/weak* judgment entails processing mostly on the potency dimension, *fast/slow* judgment requires processing largely on the activity dimension, and *active/passive* judgment involves processing on both potency and activity dimensions.¹ IAT effects should occur between *strong/weak* and *active/passive* judgments and between *active/passive* and *fast/slow* judgments, but not between *strong/weak* and *fast/slow* judgments. More specifically, for *strong/weak* and *active/passive* judgments, *strong* and *active* responses and *weak* and *passive* responses should be compatible, whereas *strong* and *passive* responses and *weak* and *active* responses should be incompatible with each other. For *active/passive* and *fast/slow* judgments, *active* and *fast* responses and *passive* and *slow* responses should be compatible, and *active* and *slow* responses and *passive* and *fast* responses should be incompatible. In contrast, for *strong/weak* and *fast/slow* judgments, no response compatibility effect should be detected.

Method

Subjects. Twenty-four students at Vanderbilt participated in this experiment for course credit. All subjects were native speakers of English. They had normal or corrected-to-normal vision. All subjects signed a consent form for participation.

Stimuli. There were 288 experimental items and 24 filler items in this experiment. The experimental items are listed in Appendix B. There were 96 experimental items for each judgment task. Most of the items came from our replication of the semantic differential. All items were selected so that they should be easier to judge in terms of their respective judgment task than the other two judgment tasks in the experiment. For every subject, experimental items for each judgment task were randomly divided into four subsets of 24 items each. Two subsets of items were presented in each block, one for each judgment task. Each item was presented only once in the experiment. The average word frequency for the experimental items was 88 per million, with an *SD* of 123 (Kučera & Francis, 1967; *ter-mite* does not have a corresponding entry in the Kučera and Francis norms). Word frequencies did not differ between items of different judgments. The item length varied between 3 and 10 letters.

Procedure. The same procedure as in Experiment 2 was followed.

Results and Discussion

Error rates were not calculated as in Experiment 2 because, for many items, there was a fairly even split across

subjects in terms of how the item was judged. There were no strong normative answers for these items with respect to the judgments. Instead, responses with RTs longer than 5,000 msec were considered errors and were excluded from the analysis. This eliminated 1.3% of the data collected. One subject (the 24th subject, literally) seemed to have adopted a very different strategy from the rest of the subjects. This subject showed IAT effects of -181 and 672 msec in the *strong/weak* and *active/passive* conditions and *strong/weak* and *fast/slow* conditions, respectively, which were -1.48 and 4.24 *SDs* away from average IAT effects in these conditions for all other subjects. The following analysis therefore did not include data from this subject.²

Figure 6 displayed mean RTs in various conditions of the experiment. For the associated judgment pair of *strong/weak* and *active/passive* and that of *active/passive* and *fast/slow*, the subjects were 117 and 179 msec slower with the hypothesized incompatible response mapping than with the hypothesized compatible response mapping. For the orthogonal judgment pair of *strong/weak* and *fast/slow*, there was only a 26-msec difference between the two response mapping conditions. A 3 (judgment pair) \times 2 (response mapping) repeated measures ANOVA confirmed that there was a main effect of response mapping [$F(1,22) = 29.10$, $\eta_p^2 = .57$], as well as an interaction between judgment pair and response mapping [$F(2,44) = 3.63$, $\eta_p^2 = .14$]. Planned comparisons suggested that the interaction was due to significant effects of response mapping for the associated judgment pair of *strong/weak* and *active/passive* and for that of *active/passive* and *fast/slow* [$F(1,44) = 8.69$ and 20.35 , $\eta_p^2 = .16$ and $.32$, respectively] and a null effect of response mapping for the orthogonal pair *strong/weak* and *fast/slow* [$F(1,44) = 0.42$, n.s.]. Paired-sample *t* tests showed that the effect of response mapping did not differ between the two judgments within pairs.

Results from the item analysis paralleled those from the subject analysis. For the associated judgment pair of *strong/weak* and *active/passive* and that of *active/passive* and *fast/slow*, items were responded to significantly slower in the incompatible response mapping condition than in

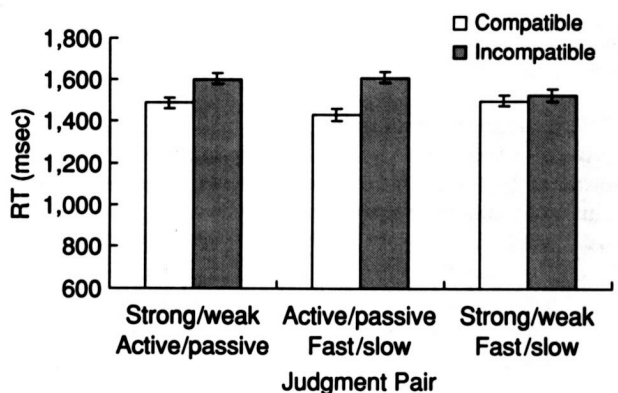


Figure 6. Reaction times (RTs) in Experiment 3, shown with data from an outlier subject removed. Error bars stand for standard errors from the ANOVA.

the compatible response mapping condition [$t_s(190) = 3.61$ and 6.00 , respectively]. For the orthogonal judgment pair of *strong/weak* and *fast/slow*, items were responded to with similar latencies in the two response mapping conditions [$t(190) = 0.68$, n.s.].

In Experiment 3, we found relatively large IAT effects between associated semantic judgments, as well as a lack of IAT effect between orthogonal semantic judgments, with association and orthogonality defined by the semantic differential approach. Experiment 3 replicated the previous two experiments with a different set of judgment tasks, verifying associations in semantic differential space as driving forces for IAT effects between semantic judgments.

GENERAL DISCUSSION

In the present study, we tested the semantic differential as a model for similarity relations among semantic judgment tasks. The hypothesis was that semantic judgment tasks should be similar if they require processing on the same dimension(s) and dissimilar if they entail processing on orthogonal dimensions in semantic space. We determined which semantic dimensions a judgment task would tap into through the corresponding semantic scale's loadings on semantic factors (Osgood et al., 1957). We used the IAT effect as a measure of strength of association (Greenwald et al., 1998). In all three experiments, we found significant IAT effects between semantic judgments when the semantic scales were relatively heavily loaded on common semantic factor(s); meanwhile, we observed no reliable IAT effects between semantic judgments when the semantic scales were loaded primarily on orthogonal semantic factors. Furthermore, all IAT effects were in the direction predicted by the semantic differential. Responses were made faster when positive and negative ends of two semantic scales were mapped to separate keys than when positive and negative ends of two semantic scales shared response keys. These results supported the semantic differential as a functional model for relations between semantic judgment tasks, which provides a well-defined basis for studying conceptual task processes in various experimental paradigms, including task switching, dual task, and transfer paradigms.

IAT and Repetition Priming Approaches to the Semantic Differential Account

In Experiment 3, we found relatively large IAT effects between the associated *strong/weak* and *active/passive* judgments and *active/passive* and *fast/slow* judgments, and we found no IAT effect between the orthogonal *strong/weak* and *fast/slow* judgments. Xiong et al. (2003) found repetition priming between *strong/weak* and *active/passive* judgments and between *active/passive* and *fast/slow* judgments, but they found no repetition priming between *strong/weak* and *fast/slow* judgments. Both results from the IAT and results from the repetition priming procedure supported the semantic differential account of similarity between semantic judgments. However, results from the IAT and repetition priming procedures diverged regarding

good/bad and *large/small* judgments. Data from our lab suggested significant repetition priming between *good/bad* and *large/small* judgments, which according to the semantic differential account, involve processing on orthogonal semantic dimensions. In the present study, the absence of IAT effect between *good/bad* and *large/small* judgments indicated that *good/bad* and *large/small* judgments should be orthogonal to each other and that repetition priming between *good/bad* and *large/small* judgments should be due to factors other than task similarity (Franks et al., 2000). Association in semantic differential space may predict transfer between semantic judgment tasks, but observed facilitation between two task performances may not imply similarity between task processes.

The IAT procedure constitutes a valid alternative method for studying associations between task processes. In addition, the IAT procedure explicates the positive or negative direction of association between semantic judgments, which has not been shown with the repetition priming paradigm. Xiong et al. (2003, Experiment 18) found the same amount of repetition priming between compatible and incompatible response mapping conditions for *pleasant/unpleasant* and *valuable/worthless* judgments. However, only one judgment was performed in each block in the Xiong et al. study, whereas in the present study, two semantic judgments were mixed within a block. Duncan (1979) argued that when two tasks were presented in a divided attention paradigm, subjects had to resolve the response mapping between tasks on a trial-to-trial basis, which they did not need to do when there was only one task or one mapping in the block. This emergent property of the combined tasks explains the discrepancy between the present results and the results of the Xiong et al. study. Also, we suggest that the IAT effect is a working memory phenomenon, mediated by activation (Frank, Loughry, & O'Reilly, 2001; Rougier & O'Reilly, 2002), whereas repetition priming can be attributed to retrieval of stored instances from long-term memory (Logan, 1988, 1990) or connection-weight-based incremental learning of the cortical system (McClelland, McNaughton, & O'Reilly, 1995; McClelland & Rumelhart, 1985; Nadel, 1992; Squire, 1992). Although both are based on the semantic differential framework, Xiong et al.'s study and the present study tapped into different forms of memory. The results are complementary rather than conflicting.

Sources of IAT Effects

Although the present study was not designed to resolve continuing issues concerning the mechanisms that underlie IAT effects, the present findings do provide some pertinent new information that could help in designing future research to clarify these mechanisms. Researchers have debated whether IAT effects actually reflect associations between target concepts or merely associations between attributes of the stimuli (De Houwer, 2001; Steffens & Plewe, 2001). In a typical IAT, the target concept *flower* was *pleasant*, and all stimulus items for the *flower* category were also *pleasant*. Associations between target concepts were confounded with associations between

attributes of the stimuli. Using *British/foreign* as target concepts, De Houwer designed an IAT that included equal numbers of positively and negatively valenced stimulus items for both *British* and *foreign* categories. IAT effect was found to be contingent on the target concepts instead of the valence of the stimuli. In contrast, Steffens and Plewe reported substantial influence of associations between stimulus items in *female/male* IATs. IAT effect was much smaller when *pleasant* stimuli were stereotypically *male* and *unpleasant* stimuli were stereotypically *female* than when *pleasant* stimuli were stereotypically *female* and *unpleasant* stimuli were stereotypically *male*. In Experiment 1 of the present study, all stimulus items were selected to be clearly positive or negative in terms of the relevant judgment and were more or less neutral with respect to the irrelevant judgments. This minimized the correlation between the *good/bad* and *hard/soft* properties of the stimuli and the correlation between the *hard/soft* and *large/small* properties of the stimuli. In particular, the correlation between the *hard/soft* and *large/small* properties of the stimulus items was .06, according to ratings from our replication of the semantic differential. No IAT effect would have been found between *hard/soft* and *large/small* judgments had associations between stimulus items been responsible for IAT effects. Thus, the present results support the importance of associations between target concepts in IAT effects.

A second potential contribution to IAT research involves an interesting methodological characteristic related to the nature of semantic differential space. The IAT research paradigm shares a family resemblance with the associative priming paradigm (e.g., doctor–nurse; Carr & Dagenbach, 1990; Neely, 1976) in that both are concerned with immediate semantic processing of stimuli. A fruitful distinction that aided theoretical understanding of associative priming involved the contrast between facilitation and inhibition effects. Assessment of these contrasting effects necessitated the choice of a neutral baseline.

Although what constituted a neutral baseline condition was never definitively resolved, the use of relatively neutral conditions was an important aspect of the associative priming paradigm.

Semantic differential space provides a theoretically motivated definition of what constitutes a neutral baseline condition in IAT. Orthogonal scales are unrelated to one another—that is, are neutral with respect to one another. RTs to semantic judgments in orthogonal conditions could be treated as neutral baselines. Data in the present experiments were reanalyzed from this perspective. The results indicated that both facilitation and inhibition contributed to IAT effects. Compatible versus incompatible mappings were arbitrary for orthogonal judgment pairs; therefore, not surprisingly, RTs for semantic judgments did not differ statistically between response mapping conditions in orthogonal pairs. Hence, the baseline RTs were collapsed across response mapping conditions. Figure 7 displays RTs to *good/bad*, *large/small*, *strong/weak*, and *fast/slow* judgments when they were in compatible response mapping, orthogonal, and incompatible response mapping conditions in each experiment. In Experiments 1 and 3, there were significant linear trends across the three conditions [*good/bad* and *large/small* judgments in Experiment 1, $F_s(1,23) = 22.58$ and 27.88 , $\eta_p^2s = .50$ and $.55$, respectively; *strong/weak* and *fast/slow* judgments in Experiment 3, $F_s(1,22) = 5.40$ and 10.42 , $\eta_p^2s = .20$ and $.32$, respectively]. In Experiment 2, the linear trends were less obvious [*good/bad* and *large/small* judgments, $F_s(1,23) = 3.89$ and 2.08 , $ps = .06$ and $.16$, $\eta_p^2s = .14$ and $.08$, respectively]. *Good/bad* judgment exhibited primarily facilitation with compatible response mapping, whereas *large/small* judgment demonstrated mainly inhibition with incompatible response mapping. In all IATs, only a single judgment was required in each trial. The facilitation and inhibition that arose from a judgment being paired with an associated judgment, relative to an orthogonal judgment, indicated that the subjects nonetheless processed the two

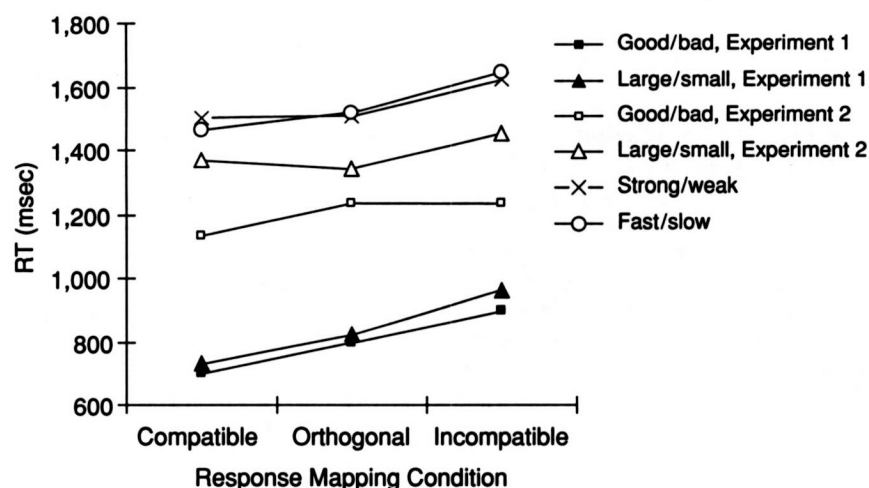


Figure 7. Facilitation and inhibition in IAT effects.

associated judgments simultaneously within a trial. Future research assessing such facilitation and inhibition processes should prove theoretically fruitful for clarifying the mechanisms underlying the IAT effect.

Summary

The present study supported the semantic differential as a functional model of semantic task processing. The results cautioned against using repetition priming as the sole measure of similarity between task processes. In addition, IAT with orthogonal judgments could constitute an ideal baseline condition for investigating IAT effects.

REFERENCES

- ARRINGTON, C. M., ALTMANN, E. M., & CARR, T. H. (2003). Tasks of a feather flock together: Similarity effects in task switching. *Memory & Cognition*, 31, 781-789.
- BENTLER, P. M., & LAVOIE, A. (1972). An extension of semantic space. *Journal of Verbal Learning & Verbal Behavior*, 11, 174-182.
- CARR, T. H., & DAGENBACH, D. (1990). Semantic priming and repetition priming from masked words: Evidence for a center-surround attentional mechanism in perceptual recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 16, 341-350.
- DE HOUWER, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, 37, 443-451.
- DE HOUWER, J. (2002). The Implicit Association Test as a tool for studying dysfunctional associations in psychopathology: Strengths and limitations. *Journal of Behavior Therapy & Experimental Psychiatry*, 33, 115-133.
- DUNCAN, J. (1979). Divided attention: The whole is more than the sum of its parts. *Journal of Experimental Psychology: Human Perception & Performance*, 5, 216-228.
- FRANK, M. J., LOUGHRY, B., & O'REILLY, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, & Behavioral Neuroscience*, 1, 137-160.
- FRANKS, J. J., BILBREY, C. W., LIEN, K. G., & McNAMARA, T. P. (2000). Transfer-appropriate processing (TAP) and repetition priming. *Memory & Cognition*, 28, 1140-1151.
- GREENWALD, A. G., MCGHEE, D. E., & SCHWARTZ, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality & Social Psychology*, 74, 1464-1480.
- GREENWALD, A. G., & NOSEK, B. A. (2001). Health of the implicit association test at age 3. *Zeitschrift für Experimentelle Psychologie*, 48, 85-93.
- KARPINSKI, A., & HILTON, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality & Social Psychology*, 81, 774-788.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LOGAN, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492-527.
- LOGAN, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, 22, 1-35.
- MCCLELLAND, J. L., McNAUGHTON, B. L., & O'REILLY, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419-457.
- MCCLELLAND, J. L., & RUMELHART, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114, 159-188.
- MEDIN, D. L., GOLDSTONE, R. L., & GENTNER, D. (1993). Respects for similarity. *Psychological Review*, 100, 254-278.
- MURPHY, G. L., & MEDIN, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289-316.
- NADEL, L. (1992). Multiple memory systems: What and why. *Journal of Cognitive Neuroscience*, 4, 179-188.
- NAVON, D., & GOPHER, D. (1979). On the economy of the human-processing system. *Psychological Review*, 86, 214-255.
- NEELY, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition*, 4, 648-654.
- OSGOOD, C. E. (1949). The similarity paradox in human learning: A resolution. *Psychological Review*, 56, 132-143.
- OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- ROEDIGER, H. L., III, & McDERMOTT, K. B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 8, pp. 63-131). Amsterdam: Elsevier.
- ROEDIGER, H. L., III, WELDON, M. S., & CHALLIS, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3-41). Hillsdale, NJ: Erlbaum.
- ROUGIER, N. P., & O'REILLY, R. C. (2002). Learning representations in a gated prefrontal cortex model of dynamic task switching. *Cognitive Science*, 26, 503-520.
- SCHACTER, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 13, 501-518.
- SQUIRE, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, 4, 232-243.
- STEFFENS, M. C., & PLEWE, I. (2001). Items' cross-category associations as a confounding factor in the Implicit Association Test. *Zeitschrift für Experimentelle Psychologie*, 48, 123-134.
- THOMPSON-SCHILL, S. L., & GABRIELI, J. D. E. (1999). Priming of visual and functional knowledge on a semantic classification task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 25, 41-53.
- TVERSKY, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.
- TZENG, O. C. S. (1975). Reliability and validity of semantic differential E-P-A markers for an American English representative sample. *Psychological Reports*, 37, 292.
- VAIDYA, C. J., GABRIELI, J. D. E., KEANE, M. M., MONTI, L. A., GUTIERREZ-RIVAS, H., & ZARELLA, M. M. (1997). Evidence for multiple mechanisms of conceptual priming on implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 23, 1324-1343.
- VRIEZEN, E. R., MOSCOVITCH, M., & BELLOS, S. A. (1995). Priming effects in semantic classification tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 933-946.
- XIONG, M. J., FRANKS, J. J., & LOGAN, G. D. (2003). Repetition priming mediated by task similarity in semantic classification. *Memory & Cognition*, 31, 1009-1020.

NOTES

1. Although *active/passive* was loaded almost exclusively on the activity factor according to Analysis I, later analyses by Osgood et al. (1957, Analysis II and the Thesaurus study) showed that *active/passive* was highly loaded on both potency and activity factors.

2. Analysis with data from the 24th subject included did not differ qualitatively from analysis without his/her data, although the interaction between judgment pair and response mapping did not reach significance level [$F(2,46) = 2.09, p = .14, \eta_p^2 = .08$]. When data from Experiments 2 and 3 were combined with experiment as the between-subjects variable and judgment pair and response mapping as within-subjects variables, the interaction between judgment pair and response mapping was statistically reliable [$F(2,92) = 3.23, \eta_p^2 = .07$].

(Continued on next page)

APPENDIX A
Experimental Items in Experiments 1 and 2

<i>good/bad</i>		<i>hard/soft</i>		<i>large/small</i>	
Positive	Negative	Positive	Negative	Positive	Negative
Experiment 1					
friend	gossip	chain	blanket	city	candle
music	pain	hammer	cloud	crowd	lemon
sleep	problem	steel	milk	field	nucleus
star	sorrow	wall	water	lake	sand
Experiment 2					
angel	acid	cabin	apron	athlete	ankle
bagel	alarm	cage	bath	bear	bait
book	burden	cave	blanket	canyon	ball
breeze	crime	chain	blood	chief	bird
bridge	critic	desk	bread	church	bowl
candy	death	door	cloud	city	candle
dawn	debt	factory	coat	college	card
dollar	decay	fence	cream	crowd	cell
evening	dilemma	gate	dream	desire	coin
family	divorce	ground	feather	doubt	crumb
farmer	failure	hail	flower	dragon	daisy
fortune	fear	hammer	girl	field	duck
friend	fever	house	hair	fire	fork
game	fight	judge	hand	forest	grain
garden	flood	king	heart	history	leak
glory	fury	locker	honey	lake	lemon
husband	germ	metal	melody	land	mouse
kiss	gossip	mirror	milk	lawn	needle
love	illness	motor	mind	life	nucleus
music	itch	paddle	mouth	moon	paper
offer	junk	police	napkin	nation	pencil
order	loss	rake	peace	nature	penny
oxygen	mold	road	perfume	ocean	plate
picture	noise	rule	puppy	orbit	pulp
profit	pain	ship	quilt	palace	salt
promise	pimple	sign	rabbit	plane	sand
sale	problem	statue	rope	river	stain
star	skull	steel	silk	scene	staple
start	slave	stone	soul	sleep	stream
teacher	sorrow	train	spirit	space	sweat
voice	trouble	tree	spring	winter	termite
wish	worry	wall	water	world	weed

APPENDIX B
Experimental Items in Experiment 3

<i>strong/weak</i>			<i>active/passive</i>			<i>fast/slow</i>		
accent	hand	perfume	acid	fence	penny	airship	failure	mold
accuser	honey	pretence	alarm	fight	personnel	ankle	field	money
apron	house	price	artist	flood	picture	ball	fire	oven
athlete	husband	principle	assistance	frost	pimple	bath	flyer	paradise
barn	idol	prosperity	bait	game	plate	bet	frame	pedal
belief	illiteracy	punishment	bear	germ	poet	bite	frog	plug
bird	impetus	resentment	beehive	girl	police	blade	fudge	pouch
boat	impunity	road	blanket	ground	problem	boot	gossip	raid
bridge	inaction	rule	blood	heat	promise	breakfast	groan	relic
cage	influence	scorn	book	hill	pulp	bunch	guess	repair
chain	inhabitant	selfishness	bowl	identification	puppy	burn	hide	river
chief	insistence	sentiment	brain	inhibition	puzzle	cache	horror	scratch
church	insolence	ship	bread	initiation	race	candle	illness	scream
company	judge	skull	cabin	itch	rake	cavern	joke	sharpness
court	king	solemnity	cave	junk	receipt	cheese	juice	shiver
cross	lake	soul	cell	kiss	sale	chicken	jump	shutter
death	land	spring	city	leak	sand	chop	lamp	sprint
debt	lemon	star	cloud	locker	scene	college	lane	steak
doubt	life	steel	concurrence	loss	slave	compression	lap	tape
dragon	love	stream	convict	lymph	sleep	crock	latch	throw
family	metal	street	crowd	market	sorrow	deck	lightning	touch
fear	mouse	sweat	crumb	meeting	space	desire	liquid	toy
feather	napkin	threat	decoy	moon	spirit	dew	lock	train
flower	nation	tree	desk	motor	start	diet	lunch	trick
fork	nature	trouble	discharge	mouth	stone	disc	lunge	trim
friend	neglect	tyrant	doctor	nucleus	story	ditch	luxury	twilight
garlic	news	unanimity	door	office	teacher	dock	maker	warehouse
gate	number	unrest	dream	orbit	termite	drink	mall	welt
ghost	opposition	upheaval	duty	order	voice	driver	marvel	whisper
grocery	pain	utterance	evening	paddle	wall	duck	merchant	wiggle
hair	palace	wish	factory	paint	window	dump	middle	wind
hammer	peasant	world	farmer	paper	winter	elf	midnight	wonder

(Manuscript received July 1, 2004;
revision accepted for publication June 29, 2005.)