Interfacing Relational Frame Theory with Cognitive Neuroscience: Semantic Priming, The Implicit Association Test, and Event Related Potentials

Dermot Barnes-Holmes*[1], Carmel Staunton[1], Yvonne Barnes-Holmes[1], Robert Whelan*, Ian Stewart[2], Sean Commins[1], Derek Walsh[1], Paul M. Smeets[3], and Simon Dymond[4]

[1]National University of Ireland, Maynooth, Ireland, [2]National University of Ireland, Galway, Ireland, [3]Leiden University, The Netherlands, [4]Anglia Polytechnic University, United Kingdom

Abstract

The current article argues that an important component of the research agenda for Relational Frame Theory will involve studying the functional relations that obtain between environmental events and the physiological activity that takes place inside the brain and central nervous system, with a particular focus on human language and cognition. In support of this view, five separate experiments are outlined. The first three experiments replicate and extend previous research reported by Hayes and Bisset (1998). Specifically, the research, using both reaction time and neurophysiological measures, supports the argument that there is a clear functional overlap between semantic and derived stimulus relations. Specifically, an evoked potential waveform typically associated with semantic processing (N400) is shown to be sensitive to equivalence versus non-equivalence relations. Experiments 4 and 5 indicate that these reaction time and evoked potential effects are not restricted to traditional lexical decision tasks, but can also be observed using the implicit association test. Furthermore, preliminary evidence suggests that evoked potentials might constitute a more sensitive measure of derived stimulus relations than response time. The results obtained across all five experiments support the view that the study of derived stimulus relations, combined with some of the procedures and measures of cognitive psychology and cognitive neuroscience, may provide an important inroad into the experimental analysis of semantic relations in human language.

Key words: Relational Frame Theory, cognitive neuroscience, semantic priming, implicit association test, event related potentials.

Resumen

El presente artículo sostiene que una parte importante dentro del programa de investigación de la Teoría del Marco Relacional será el estudio de las relaciones funcionales entre eventos ambientales y la actividad fisiológica que tiene lugar en le cerebro y el sistema

* The current article is dedicated to the memory of Carmel Staunton, who lost her life tragically in a road traffic accident in October 2003. Sections of the current article were published in the Irish Psychologist, in January 2004. Address all correspondence to Dermot Barnes-Holmes, Department of Psychology, National University of Ireland, Maynooth, Maynooth, Co. Kildare, Ireland. Email: Dermot.Barnes-Holmes@may.ie.

nervioso central, con un énfasis particular en el estudio del lenguaje y la cognición humanos. Para apoyar este punto de vista, se presenta un breve esbozo de cinco experimentos diferentes. Los tres primeros replican y amplían el trabajo de Hayes y Bisset (1998). Específicamente, estas investigaciones, empleando tanto medidas de tiempo de reacción como medidas neurofisiológicas, apoyan el argumento de que hay un claro solapamiento funcional entre las relaciones semánticas y las relaciones derivadas entre estímulos. Concretamente, se observa que un componente de potenciales evocados (un potencial evocado) típicamente asociado con el procesamiento semántico (N400) es sensible a las relaciones de equivalencia frente a las de no equivalencia. Los experimentos 4 y 5 muestran que estos efectos en tiempos de reacción y potenciales evocados no están limitados a las tareas tradicionales de decisión léxica, sino que también pueden ser observados cuando se emplea el test de asociación implícita. Es más, la evidencia preliminar sugiere que los potenciales evocados podrían ser una medida de relaciones derivadas más sensible que el tiempo de reacción. Los resultados obtenidos de manera general en los cinco experimentos dan apoyo a la idea de que la combinación entre el estudio de las relaciones derivadas entre estímulos y algunas de las técnicas y medidas habitualmente empleadas por la psicología cognitiva y la neurociencia cognitiva, puede constituir una importante vía de investigación para el análisis experimental de las relaciones semánticas en el lenguaje humano.
Palabras clave: Teoría del marco relacional, neurociencia cognitiva, priming semántico, test de asociación implícita, potenciales relacionados con eventos.

Behavioral psychologists, it has been argued, seek to develop a science of behavior that is independent, yet complementary to, the neurosciences (e.g., Barnes & Hampson, 1997; DiFore, Dube, Oross, Wilkinson, Deutsch, & McIlvane, 2001). As a behavioral account of human language and cognition, Relational Frame Theory (RFT) is part of this tradition (Hayes, Barnes-Holmes, & Roche, 2001). It follows, therefore, that a critical component of the RFT research agenda should involve studying the functional relations that obtain between environmental events and the physiological activity that takes place inside the brain and central nervous system, with a particular focus on human verbal behavior. Admittedly, this type of research is only in its infancy. Indeed, the first author is one of the few behavioral psychologists to have published research that has attempted to integrate the study of neural-network models with a behavioral theory of human language and cognition (e.g., Barnes & Hampson, 1997).

A logical and indeed vital extension of this earlier work would involve studying neural activity as it occurs during the performance of specific verbal or cognitive tasks. One ideal methodology for research in this area involves measuring what have been called event-related potentials (ERPs). These measures are averaged segments of electroencephalograms (EEGs) that are time-locked to a specific type of stimulus. The waveforms, or components, that emerge following the averaging procedure provide a measure of the brain activity that is functionally related to the time-locked stimulus. Event related potentials therefore allow the researcher to examine neural events that occur between the onset of a stimulus (e.g., a word on a computer screen) and an overt response (e.g., a key press). Although it is difficult to identify the specific location of the neural activity that produces these waveforms, ERPs can provide a measure of the

summed activity of the brain with timing in the order of milliseconds. As argued by Barnes and Hampson (1997), measuring neural events that occur in this short temporal gap will be absolutely critical in developing a more complete understanding of language and cognition from a behavioral perspective.

Of course, calling for the study of such neural events is relatively straightforward. It is quite another matter to undertake the painstaking and difficult work involved in developing and refining the necessary experimental procedures, and gathering the relevant data sets, that are necessary to uncover the nature of the neural activity that is functionally related to human verbal behavior. In the current article, we will outline a recent program of RFT research that is aiming to address this issue.

In particular, our research program constitutes one of the first steps toward interfacing the behavioral and cognitive neuroscience approaches to semantic processing in natural language.

Relational Frame Theory and the Behavioral Analysis of Human Language and Cognition

One of the core assumptions of RFT is that the behavioral units of human language and thought may be defined in terms of derived stimulus relations and relational networks (Barnes & Holmes, 1991; Hayes, et al., 2001). Perhaps the simplest example of a derived stimulus relation is the equivalence relation, which some have argued provides the basis for semantic or symbolic meaning in natural language (e.g., Sidman, 1986, 1994). Equivalence relations are often examined in the behavioral laboratory through the use of a matching-to-sample (MTS) procedure. This procedure involves training participants to match abstract stimuli to each other and then presenting a series of test or probe trials to determine if predictable, but untrained, matching performances emerge.

In a typical computerized MTS trial, a participant might be presented with the nonsense word CUG as a sample stimulus and ZID as one of two comparison stimuli. If the participant chooses ZID, the word "Correct" is presented; if the other comparison is chosen "Wrong" appears. On another trial, the word ZID may be presented as a sample stimulus along with DAX as one of two comparisons; choosing DAX produces "Correct" on the screen and choosing the other comparison produces "Wrong". This training may be represented as follows:

CUG -> ZID and ZID -> DAX

In order to test for an equivalence relation, a number of test or probe trials are presented *in the absence of any corrective feedback*. For example, ZID may be presented as a sample with CUG as one of the comparisons. If the participant reliably chooses CUG given ZID this provides evidence for what is called symmetry. In effect, training CUG -> ZID leads to a derived symmetrical relational response (i.e., ZID -> CUG). If the participant also shows DAX -> ZID symmetry and what is called transitivity (i.e., CUG -> DAX) and combined symmetry and transitivity (i.e., DAX -> CUG) the three stimuli are said to participate in an equivalence relation. (Parenthetically, combined

symmetry and transitivity is sometimes referred to as an equivalence relation; see Sidman, 1990). The foregoing training and test trials may be represented as follows: Train A->B and B->C, and test B->A, C->B (symmetry), A->C (transitivity), and C->A (equivalence).

It should be noted, that the foregoing example describes only one of the available designs used for training and testing equivalence relations. For example, numerous studies have trained A->B and A->C relations and then tested for B->A and C->A symmetry relations and the two combined symmetry and transitivity relations; B->C and C->B. The design described previously is referred to as a linear protocol, whereas the latter design is referred to as a one-to-may or sample-as-node protocol. Furthermore, equivalence relations may contain more than three members. In the some of the work to be described subsequently, for example, four-member equivalence relations were trained and tested using a linear protocol.

Derived Stimulus Relations and Human Language

There are many findings supporting the connection between equivalence relations and human language (see Horne & Lowe, 1996, for a review). First, evidence suggests that verbal abilities and the capacity to derive stimulus relations co-vary (e.g., Barnes, McCullagh, & Keenan, 1990; Devany, Hayes, & Nelson, 1986), although the source of that co-variation is still at issue (e.g., Leslie & Blackman, 2000). Second, it is known that derived stimulus relations develop in very early childhood (Lipkens, Hayes, & Hayes, 1993; Luciano, Barnes-Holmes, & Barnes-Holmes, 2001) and can be delayed by a lack of exposure to verbal training (Barnes, et al., 1990). Third, derived stimulus relations are at least very difficult to produce and are arguably absent in nonhumans (García & Benjumea, 2001; Hayes & Hayes, 1992; Dugdale & Lowe, 2000; but see Shusterman & Kastak, 1993). Fourth, equivalence, exclusion, and similar procedures have often been used as a means of establishing novel verbal performances (de Rose, de Souza, Rossito, & de Rose, 1988; Sidman, 1971). Finally, one recent study has shown that brain activation patterns produced during the formation of equivalence relations (recorded using fMRI) resemble those involved in semantic processing underlying language (Dickins, Singh, Roberts, Burns, Downes, Jimmieson, & Bentall, 2001; see also DiFore, et al., 2001).

Apart from the growing body of empirical evidence that appears to support a connection between derived relations and human language, a number of behavioral researchers have also argued that traditional network theories of verbal or semantic meaning (e.g., Anderson, 1976, 1983; Collins & Loftus, 1975; McClelland & Rumelhart, 1988) share similarities with the concept of derived stimulus relations (Barnes & Hampson, 1993; Cullinan, Barnes, Hampson, & Lyddy, 1994; Fields, 1987; Hayes & Hayes, 1992; Reese, 1991). At the present time, however, the available evidence to support the argument that semantic networks and derived stimulus relations possess similar properties is somewhat limited. In fact, only two published studies appear to speak directly to this particular issue (Dickins, et al., 2001; Hayes & Bisset, 1998).

Equivalence as a Behavioral Model of Semantic Relations: The Priming Hypothesis

If derived stimulus relations serve as a useful working model of semantic meaning, then the pattern of findings that have been demonstrated using semantic stimuli should also be found when using stimuli from equivalence or other derived relations (Branch, 1994). As a first step towards testing this basic postulate, Hayes and Bisset (1998) employed a semantic priming procedure (i.e., a lexical decision task) to examine the semantic-like properties of laboratory-induced equivalence relations. In effect, their study was designed to test the argument that one of the key measures of language and thought processes typically employed within mainstream cognitive psychology should also be sensitive to derived stimulus relations in the context of episodic and mediated priming.

The prototypical priming effect is shown when a participant more rapidly recognises a word as a word when it is preceded by a related rather than an unrelated word. That is, if the two words presented in a lexical decision task are semantically related (e.g., tiger-lion) the participants' reaction times (RTs) are significantly shorter than if the words are semantically unrelated (e.g., tiger-house). The priming literature includes many variants such as semantic, associative, mediated, and episodic priming as well as numerous experimental preparations utilized to demonstrate priming, such as lexical decision and pronunciation tasks (see Neely, 1991, for a review).

If equivalence relations provide a valid behavioral model of semantic relations, then equivalence should also demonstrate priming effects (Hayes & Bisset, 1998). In order to test this suggestion, Hayes and Bisset sought to determine if priming in a lexical decision task occurs for previously trained and tested equivalence relations. Participants were first exposed to a computerized one-to-many protocol in which three, three-member equivalence classes were established using word-like nonsense words (subjects were told that the words were from a foreign language). At the end of this part of the experiment, therefore, each participant had been trained in six MTS trial-types (e.g., A1-B1 & A1-C1) and had successfully demonstrated the formation of three equivalence relations (e.g., B1-C1 & C1-B1). In the next part of the study, participants were exposed to a lexical decision task using the nonsense words employed in the equivalence training and testing. Previously unseen nonsense words were also used on some trials (this is a typical control procedure in studies of semantic priming). Subjects were asked to press a "YES" key if both words were from any of the previously learned equivalence relations, and to press a "NO" key if one or both words were previously unseen (feedback was given on each trial for correct and incorrect responses). In fact, there were seven conditions in the lexical decision task, but the most important comparison was between those trial-types in which the words were from the same equivalence relations (e.g., B1-C1) versus those trial-types in which they were not (e.g., B1-C3). Hayes and Bisset found that mean reaction times to equivalently related word pairs was significantly faster than mean reaction times to non-equivalently related word pairs. In effect, the equivalence relations appeared to generate priming effects not unlike those typically found when real words are used in cognitive research (e.g., Balota & Lorch, 1986; de Groot, 1983; McNamara & Altarriba, 1988; Meyer & Schvaneveldt, 1971).

These data therefore supported the basic postulate that derived relational responding provides a working model of semantic relations.

Limitations to the Hayes and Bisset Study

Although the research reported by Hayes and Bisset (1998) provided very clear priming effects, certain features of their experimental work limits the extent to which strong conclusions may be drawn from the data. Hayes and Bisset (1998) employed the two-word lexical decision task in which the participant is required to respond to both stimuli (i.e., YES if both are real words and NO if one or both stimuli are nonsense words). By far the most common procedure in modern priming studies is the single-word priming paradigm (see Neely, 1991). The typical procedure involves presenting a prime for a very brief period (e.g., 500 ms) and then shortly thereafter (e.g., 200 ms) presenting a target word. The participant is required to respond YES if the target is a real word and NO if it is a nonsense word. In effect, the participant responds only to the target, not to the prime (hence the name *single-word priming*). Given that reliable priming effects have been reported across numerous studies using the single-word paradigm (see Neely, Keefe, & Ross, 1989), it seems important to replicate Hayes and Bisset's data with this modern procedure if the generality of their results is to be sustained. Indeed, if priming cannot be shown using a procedure that has proved very effective within the context of natural language, this would seriously undermine the argument that derived stimulus relations provide a useful model of human language and cognition, and would therefore seriously threaten the key postulate of RFT.

Another possible limitation to the Hayes and Bisset (1998) study concerns the fact that they presented participants with corrective feedback for correct and incorrect responses during the lexical decision task. Consequently, it is difficult to separate out the effects of the feedback that occurred during the MTS training from those that occurred during the priming procedure. Some priming studies in the cognitive literature have demonstrated semantic priming in the absence of differential feedback (e.g., Hill, Strube, Roesch-Ely, & Weisbrod, 2002; Holcomb & Anderson, 1993; Weisbrod, Kiefer, Winkler, Maier, Hill, Roesch-Ely, & Spitzer, 1999), and thus it seems important also to replicate this effect with equivalence-based priming if the derived stimulus relations model of semantic meaning is to be upheld. In the next section of the article, we will outline the results from three experiments that were designed to address this issue.

Testing the Priming Hypothesis

Experiment 1. The first experiment in the series involved training and testing university undergraduates in the necessary conditional discriminations for the formation of two 4-member equivalence relations (training A1-B1, B1-C1, & C1-D1 allows for the derivation of C1-A1, D1-B1, & D1-A1; see Table 1 for a full list of the trained and tested relations). As in the Hayes and Bisset study, subjects were told that the nonsense words employed in the MTS training and testing were from a foreign language, and they had to learn how to match them. This training and testing was then followed by

Table 1. A Schematic Representation of the Trained Conditional Discriminations and Tested Equivalence Relations (Experiments 1-3).

Sample	Correct Comparison	Incorrect Comparison
Trained Conditional Discriminations		
A1	B1	B2
B1	C1	C2
C1	D1	D2
A2	B2	B1
B2	C2	C1
C2	D2	D1
Tested Equivalence Relations		
D1	A1	A2
D1	B1	B2
C1	A1	A2
D2	A2	A1
D2	B2	B1
C2	C2	C1

exposure to a lexical decision task designed to test for priming effects among the stimuli participating in the equivalence relations (see Table 2 for a complete list of the priming test trials). Like the Hayes and Bisset study, the lexical decision task included trials that presented primes and targets; (i) from the same equivalence relations (Class–Class trials), (ii) from different equivalence relations (Class–Nonclass), and (iii) previously unseen nonsense words (e.g., Nonsense–Class). Unlike the Hayes and Bisset study, however, a single-word priming paradigm was employed, and no feedback was provided during the lexical decision task. If priming effects are observed among directly and indirectly related members of the equivalence classes this would provide strong evidence for priming among derived stimulus relations, and therefore support the assumption that such relations provide a valid behavioral model of semantic meaning in natural language.

The data obtained from this first experiment showed that priming effects, as measured by reaction times, can be obtained through derived stimulus relations, whether directly or indirectly related. The stimulus pairs from the same equivalence relations primed each other more rapidly than stimuli from different equivalence relations or when pairs contained one or two previously unseen stimuli (see Figure 1). There were no significant differences in priming between any of the within-equivalence class comparisons or among any of the comparisons between the conditions that contained non-equivalent or previously unseen stimuli. In summary, the current data replicate the RT effects reported by Hayes and Bisset (1998).

The priming effects, as measured by RTs, observed in the first experiment replicated the findings of Hayes and Bisset (1998), in that priming was achieved without the

Table 2. A Schematic Representation of the 48 Trial-Types Presented During the Lexical Decision Procedure (Pm= Prime. Tg= Target. Rp= Correct Response. N= Previously Unseen Nonsense Word).

Class – Class			Class – Nonclass			Class – Nonsense			Nonsense – Class			Nonsense –Nonsense		
Pm	**Tg**	**Rp**	**Pm**	**Tg**	**Rp**	**Pm**	**Tg**	**Rp**	**Pm**	**Tg**	**Rp**	**Pm**	**Tg**	**Rp**
Directly Trained														
A1	B1	Yes	B1	A2	Yes	A1	N1	No	N1	A1	Yes	N1	N5	No
B1	C1	Yes	A1	C2	Yes	B1	N2	No	N2	B1	Yes	N2	N6	No
C1	D1	Yes	D1	B2	Yes	C1	N3	No	N3	C1	Yes	N3	N7	No
A2	B2	Yes	B2	A1	Yes	A2	N5	No	N5	A2	Yes	N5	N1	No
B2	C2	Yes	A2	C1	Yes	B2	N6	No	N6	B2	Yes	N6	N2	No
C2	D2	Yes	D2	B1	Yes	C2	N7	No	N7	C2	Yes	N7	N3	No
Symmetry														
B1	A1	Yes												
C1	B1	Yes												
D1	C1	Yes												
B2	A2	Yes												
C2	B2	Yes												
D2	C2	Yes												
Transitivity														
A1	C1	Yes												
B1	D1	Yes												
A2	C2	Yes												
B2	D2	Yes												
A1	D1	Yes												
A2	D2	Yes												
Equivalence														
C1	A1	Yes												
D1	B1	Yes												
C2	A2	Yes												
D2	B2	Yes												
D1	A1	Yes												
D2	A2	Yes												

benefit of learned semantic context (stimuli were non-words without a pre-experimental history) and also sometimes without direct association (i.e., when primes and targets were related to each other via transitive or equivalence relations). The results of Experiment 1 therefore support the view that derived stimulus relations act like semantic relations, to the extent that priming is a semantic process. Furthermore, insofar as priming is an associative process, derived stimulus relations appear to act like direct associations. However, caution is required in drawing this latter conclusion. In Experiment 1, all participants were required to pass an equivalence test before exposure to the lexical-decision task, and thus the four stimuli contained within each of the two equivalence classes had been repeatedly matched (i.e., directly associated) during the test, albeit without differential reinforcement. This fact limits the extent to which the priming effects observed in Experiment 1 can be defined as *mediated* rather than *direct* priming. This issue was addressed in the second experiment in the series.

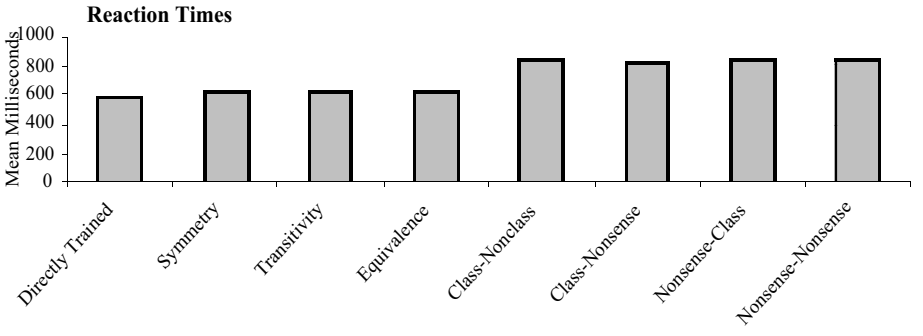

Figure 1. Priming, as measured by reaction times, for equivalent and non-equivalent stimuli in Experiment 1. Priming is indicated by lower scores.

Experiment 2. Mediated priming refers to the priming effect that is sometimes obtained when the prime and target are *indirectly* semantically related via a mediating word or concept. For example, the word *stripes* may prime recognition of *lion* based on the mediating concept *tiger*. In Experiment 1, priming was clearly demonstrated across combined symmetry and transitivity relations (e.g., D1 primed A1), and this was taken as evidence for mediated priming because the D and A stimuli were indirectly related via the B and C stimuli. However, all of the participants had successfully completed an equivalence test prior to the lexical decision task and thus the D and A stimuli had been directly related during this test (albeit in the absence of differential reinforcement). Consequently, the D-A priming observed in Experiment 1 may have simply reflected direct rather then mediated priming. Ideally, therefore, an equivalence test should not be presented until *after* the lexical decision task if unequivocal mediated priming is to be observed across indirectly related members of an equivalence relation. As an aside, Hayes and Bisset (1998) exposed their participants to an equivalence test *before* the lexical decision task, and thus their data also failed to provide strong evidence for mediated priming.

Given that mediated priming has been documented in the cognitive literature using natural language (e.g., Balota & Lorch, 1986; Weisbrod, et al., 1999), it is important that this priming effect be shown within the context of the equivalence paradigm if the RFT account of human language and cognition in terms of derived stimulus relations is to be upheld. In the second experiment, therefore, participants were given the same MTS training as that provided in Experiment 1, but were exposed to the lexical decision task *before* proceeding to the MTS equivalence test. If priming effects are observed among directly and indirectly related members of to-be-tested equivalence relations, this would provide strong evidence for both direct and mediated priming among derived stimulus relations, and therefore support the assumption that such relations provide a valid behavioral model of semantic relations in natural language.

The data obtained from Experiment 2 were divided into two sets: RTs on the

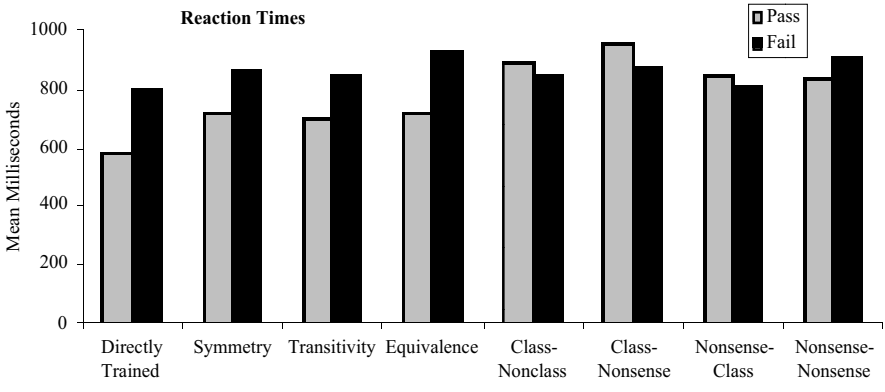

Figure 2. Priming, as measured by reaction times, for equivalent and non-equivalent stimuli for participants who passed and failed the equivalence test in Experiment 2. Priming is indicated by lower scores.

lexical decision task for participants who passed the equivalence test versus RTs for those who failed. The group who passed the equivalence test produced significantly faster reaction times to targets that were primed with same-class members than to targets that were primed with non-class members or novel stimuli. In contrast, the group who failed the equivalence test also failed to produce any evidence of priming (see Figure 2). In Experiment 2, therefore, priming effects were observed for stimulus pairs that were never directly associated (i.e., paired together), and thus the RT data in this experiment appears to provide evidence for mediated priming using derived stimulus relations.

The results of the second experiment are particularly compelling in that priming was not observed for those participants who subsequently failed the equivalence test. This indicates that training in a set of interrelated conditional discriminations is not sufficient to produce the priming effect normally observed with semantic relations in natural language. Rather, the conditional discrimination training must give rise to derived equivalence relations if semantic-like effects are to be obtained. This result certainly supports the argument that derived relations, rather than directly reinforced stimulus relations alone, provide a behavioral model of what cognitive researchers refer to as semantic processes (Barnes-Holmes, Hayes, & Dymond, 2001; Barsalou, 1999; Deacon, 1997).

The third experiment in the series constituted a further test of the RFT model of semantic relations. In this experiment, ERPs were employed as a measure of semantic processing, thereby beginning to build that important interface between RFT and cognitive neuroscience.

Experiment 3. In both Experiments 1 and 2, the most common measure of semantic

priming was used -reaction times. However, there is a substantive body of research on semantic priming that has also employed ERPs as a measure of the priming effect (e.g., Bentin, McCarthy, & Wood, 1987; Kutus & Hillyard, 1980; Weisbrod, et al., 1999). As outlined previously, ERPs are particularly well suited to studying the effects of discrete stimulus presentations on human learning (see Holcomb, 1988; Holcomb & Neville, 1991; Kutas, 1993). Specifically, the technique involves placing electrodes at specified locations on the scalp of the head, from which it is possible to record EEGs from each location (i.e., the electrical activity of groups of millions of neurons underneath each electrode). Sometimes, however, the electrical signals can be messy or noisy: it is difficult to distinguish the brain's normal background activity and the activity produced by perceiving or responding to a stimulus. To overcome this, researchers have devised the technique of averaging signals across trials. This is achieved by recording ERPs (these are sometimes referred to as *evoked potentials*), which are electrical signals time-locked to a repeatedly presented stimulus (or set of stimuli). Each EEG response to a stimulus is added and averaged to produce one clearer signal or evoked potential. The potentials are event-related because they are related to a specific stimulus event. The point of averaging is to make the effect of a stimulus on the EEG clearer; background noise is reduced and the effect of the stimulus becomes more obvious.

There are numerous waveforms associated with ERPs measures. For example, some ERPs are thought to be associated with cognitive functions such as understanding words or being able to distinguish one type of visual or auditory stimulus from another. These ERPs occur at around 300 or 400 milliseconds after the stimulus onset. The ERPs measure that is most relevant in the context of the current research is a late negative waveform, known as the N400 (see Holcomb & Anderson, 1993; Kounios & Holcomb, 1992). This waveform is typically produced when participants are asked to respond to words that are semantically unrelated. In contrast, when the words are from the same semantic categories, the N400 is greatly reduced or completely absent. In effect, the N400 has proved to be a sensitive measure of semantic relations in natural language (Holcomb & Neville, 1991). If the N400 were similarly sensitive to derived stimulus relations, this would provide additional evidence to support the derived stimulus relations' model of symbolic control.

The third experiment sought to determine if the N400 waveform would also differentiate between non-equivalent and directly trained and equivalent stimulus relations on a lexical decision task. Insofar as the N400 is more sensitive to semantic associations than reaction time (e.g., Kounios & Holcomb, 1992), demonstrating N400 sensitivity to equivalence relations would thus provide important additional evidence for the functional overlap between semantic and derived relations.

Evoked potentials were collected across eight-electrode sites, for each of the participants, while they completed the priming task. The grand average waveforms, calculated across participants, showed greater negative deflections for the non-equivalent priming trial-types than for the directly trained and equivalent trial-types, with some suggestion that the differences were greater for the left hemisphere, relative to the right (see Figure 3). The peak amplitudes of the N400 waveforms for each participant, measured between 350 and 550 ms following target onset, indicated significant effects

for electrode site and priming trial-type, with an interaction between trial-type and laterality approaching significance (i.e., p= .06). Overall, these and subsequent post-hoc analyses indicated that the negative peak amplitudes generated by the non-equivalent priming trial-types were significantly greater relative to the directly trained and equivalent priming trial-types; three sites on the left (C3, P3 & T3) showed a significant difference between the non-equivalent and both the directly trained and equivalent conditions,

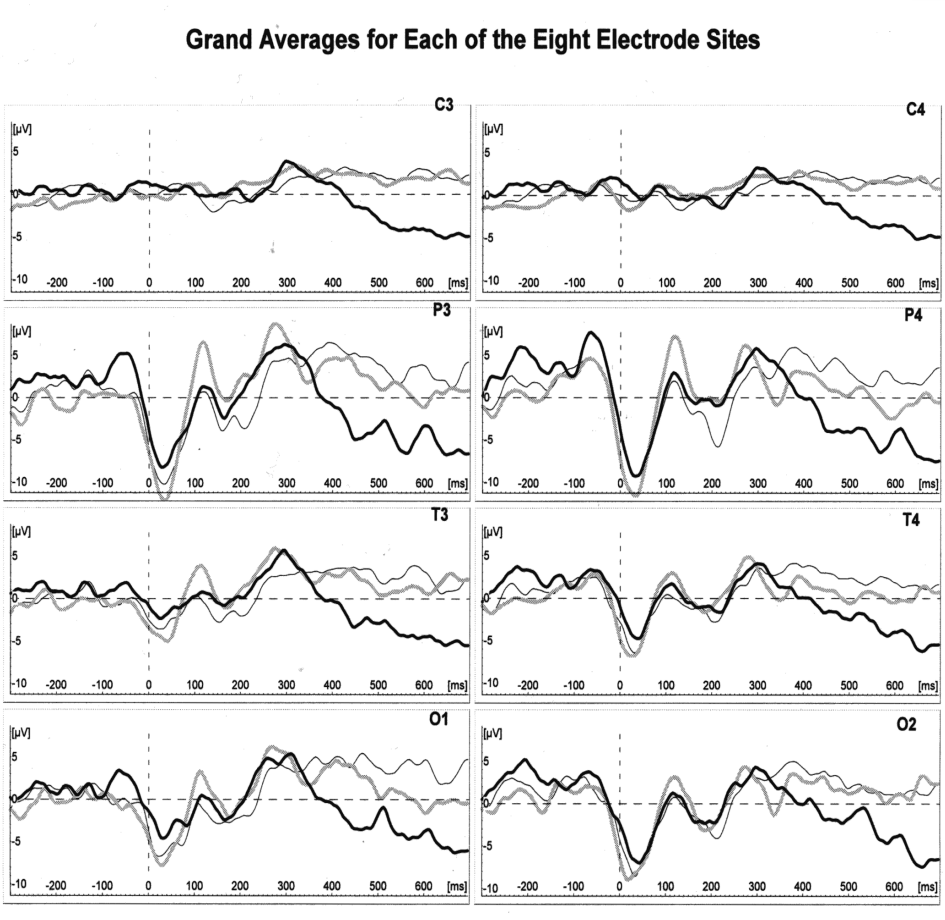

Figure 3. Grand average waveforms calculated across participants for prime-target stimulus pairs that were directly trained (thin black lines), equivalent (thick gray lines) and non-equivalent (think black lines) at electrode sites C3, C4 (top panel), P3, P4 (second from top panel), T3, T4 (third from top panel), O1 and O2 (bottom panel). Note that the prime was presented 100 ms (-100) prior to the target stimulus (0 ms). The greater negative deflections, commencing around 400 ms after target onset, for the non-equivalent prime-target pairs appear to parallel the N400 waveforms typically observed when semantically unrelated words, taken from natural language, are presented on a lexical decision task.

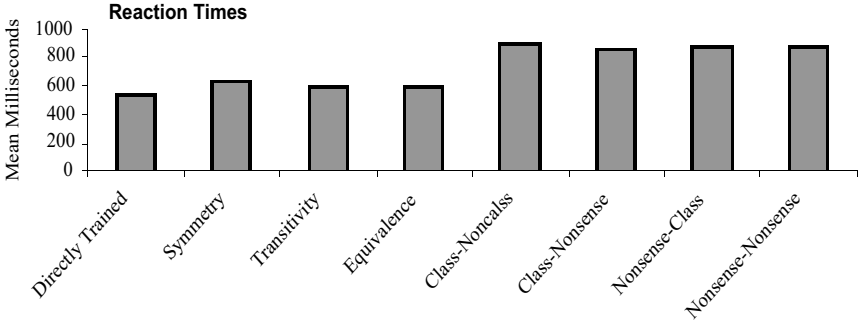

Figure 4. Priming, as measured by reaction times, for equivalent and non-equivalent stimuli in Experiment 3. Priming is indicated by lower scores in each case.

whereas one site on the left and two on the right (O1, P4 & T4, respectively) showed a significant difference between the non-equivalent and directly trained condition alone. In summary, therefore, the data from Experiment 3 demonstrated that the N400 waveform is sensitive to the difference between stimulus pairs that participate in equivalence classes versus pairs that do not, with some suggestion that this sensitivity is greater on the left than on the right. These ERPs data are therefore generally consistent with previous studies of direct and mediated priming using the N400 waveform (e.g., Weisbrod, et al., 1999).

Finally, the mean RTs in Experiment 3 (see Figure 4) showed that participants responded more rapidly to prime-target pairs that were from the same equivalence relation than to prime-target pairs that were from different equivalence relations or to those that contained one or two novel stimuli. In summary, therefore, the ERPs data collected in Experiment 3 were consistent with the reaction time measures collected across all three experiments.

Although these findings clearly support the RFT postulate that derived relations provide a behavior-analytic model of semantic relations in natural language, other relevant methodologies are available to the researcher in this area. If these methodologies also yield RFT data that is broadly consistent with the results of mainstream cognitive psychology, and cognitive neuroscience, this would further bolster the RFT concept of semantic relations. In the next part of the current article we will examine a relatively new methodology for studying such relations and consider some recent RFT data.

Relational Frame Theory, the Implicit Association Test, and Event Related Potentials

Although the lexical decision task has been used extensively in the study of semantic relations, other relevant experimental methodologies have been developed. One of the most recent of these is the so-called Implicit Association Test (IAT), which, it as has been argued, provides a more sensitive measure of semantic categories than

traditional priming procedures (e.g., Greenwald, McGhee, & Schwartz, 1998). The development of this procedure has not been without controversy, however, because it has been claimed that it can provide a measure of prejudice and other implicit attitudes that an individual would typically deny or prefer to hide (see Dasgupta, Greenwald, & Banaji, 2003). The details of this debate are not the concern of the current article, but insofar as the IAT is sensitive to semantic relations it does provide another means of testing RFT. Specifically, does the IAT produce similar effects using derived stimulus relations as are found using natural language categories?

The Implicit Association Test and RFT. The most critical components of the IAT involve two tests in which two responses (e.g., right and left key presses) are assigned to four categories, such as flowers, insects, pleasant words, and unpleasant words. In one test, for example, a participant might respond to flowers and pleasant words by pressing the right key and to insects and unpleasant words by pressing the left key. In this test, therefore, the response assignment is congruent with the typical semantic categories one might expect to find in the general population (i.e., flowers are pleasant and insects are unpleasant). Insofar as the categories assigned to each key are indeed congruent (i.e., right key = flowers/pleasant and left key = insects/unpleasant) the IAT tends to produce responding that is relatively fast because like is categorized with like.

In the other critical test of the IAT, the response assignment for flowers and insects is reversed, but the response assignment for pleasant and unpleasant words remains unchanged. Consequently, the two categories assigned to the right key (insects and pleasant) are now incongruent, as indeed are the categories assigned to the left key (flowers and unpleasant). In comparison to the first test, in which the categories are congruent, the IAT tends to produce responding that is slower because opposing categories are categorized together. Typically, of course, the order of congruent and incongruent test presentations is counterbalanced to avoid practice or negative transfer effects.

A basic RFT model of the IAT effect could involve training and testing at least two equivalence relations, and then presenting within-class probes to model the congruent categories test and across-class probes for the incongruent test. In the former case, for example, all class 1 stimuli would be assigned to the right key and all class 2 stimuli would be assigned to the left key. Thus, a participant might be instructed "to press the right key if either A1 or B1 is presented or to press the left key if either A2 or B2 is presented." In the latter case, however, pairs of class 1 and class 2 stimuli would be assigned to each key, such that a participant might be instructed "to press the right key if either A1 or B2 is presented or to press the left key if either A2 or B1 is presented." If derived equivalence relations provide a basic model of semantic categories, and the IAT effect is based, at least in part, on the juxtaposition of such categories, RFT would predict that within-class IAT probes should produce shorter average reaction times than across-class probes.

At the time of writing, the authors were unaware of any published study that had employed ERPs as a measure of IAT performance using natural language categories, and thus it was not possible to predict the ERPs waveforms based on previous IAT research. One recent study has shown that preference for White versus Black faces on the IAT is related to activation of the amygdala (Phelps, O'Connor, Cunningham,

Funayama, Gatenby, Gore, & Banji, 2000), a sub-cortical brain structure that has been implicated in emotional learning and memory. This research was focused on demonstrating that the IAT is sensitive to emotionally-laden categories, not just "cold and cognitive" associations (Dasgupta, et al., 2003). Nevertheless, the findings do not address the basic behavioral processes involved in the formation of such categories (both emotional and cognitive) and the manner in which the IAT taps into these processes. The type of RFT research described subsequently, however, might begin to provide the relevant information that we need to better understand these process issues.

Experiment 4. In this fourth experiment, participants were trained and tested in the necessary conditional discriminations for the formation of four 3-member equivalence

Table 3. A Schematic Representation of the Trained Conditional Discriminations and Tested Equivalence Relations (Experiments 4 & 5).

Sample	Correct Comparison	Incorrect Comparisons
Trained Conditional Discriminations		
A1	B1	B2, B3, B4
A1	C1	C2, C3, C4
A2	B2	B1, B3, B4
A2	C2	C1, C3, C4
A3	B3	B1, B2, B4
A3	C3	C1, C2, C4
A4	B4	B1, B2, B3
A4	C4	C1, C2, C3
Tested Equivalence Relations		
Trained Relations		
A1	B1	B2, B3, B4
A1	C1	C2, C3, C4
A2	B2	B1, B3, B4
A2	C2	C1, C3, C4
A3	B3	B1, B2, B4
A3	C3	C1, C2, C4
A4	B4	B1, B2, B3
A4	C4	C1, C2, C3
Symmetry Relations		
B1	A1	A2, A3, A4
C1	A1	A2, A3, A4
B2	A2	A1, A3, A4
C2	A2	A1, A3, A4
B3	A3	A1, A2, A4
C3	A3	A1, A2, A4
B4	A4	A1, A2, A3
C4	A4	A1, A2, A3
Equivalence Relations		
B1	C1	C2, C3, C4
C1	B1	B2, B3, B4
B2	C2	C1, C3, C4
C2	B2	B1, B3, B4
B3	C3	C1, C2, C4
C3	B3	B1, B2, B4
B4	C4	C1, C2, C3
C4	B4	B1, B2, B3

relations (see Table 3 for a full list of the trained and tested relations). Once again, subjects were told that the nonsense words employed in the MTS training and testing were from a foreign language, and they had to learn how to match them. This training and testing was then followed by exposure to an IAT that was designed to test for "implicit associations" among the stimuli participating in the equivalence relations.

Each IAT test trial involved the presentation of two instructions, one presented in the top left corner and the other in the top right corner of a computer screen, and the presentation of a single stimulus in the center of the screen. The left-side instruction was of the form, "If A or B press left" and the right-side instruction was of the form, "If X or Y press right." The A, B, X, and Y elements in these instructions refer to specific nonsense words that were employed in the previous MTS training and testing. On any given trial, one of the four nonsense words (A, B, X, or Y) was presented in the center of the screen. If the subject pressed either the left or right key the screen cleared and remained blank until the next trial (i.e., no differential feedback was provided). If the subject failed to respond within 3000 ms, the screen cleared and the words "Too

Table 4. A Schematic Representation of the 96 Trial-Types Presented During the Implicit Association Test Procedure. On Each Trial One of the Twenty-Four Instruction Sets Was Presented With One of the Four Target Stimuli.

Class – Class			Class – Nonclass		
Instruction: Press Left If	**Instruction: Press Right If**	**Target Stimuli and Correct Response**	**Instruction: Press Left If**	**Instruction: Press Right If**	**Target Stimuli and Correct Response**
Directly Trained Instruction Sets (1 – 4)			**Directly Trained Instruction Sets (13 – 16)**		
A1 or B1	A2 or B2	A1 / B1 (Left) A2 / B2 (Right)	A1 or B2	A2 or B1	A1 / B2 (Left) A2 / B1 (Right)
A1 or C1	A2 or C2	A1 / C1 (Left) A2 / C2 (Right)	A1 or C2	A2 or C1	A1 / C2 (Left) A2 / C1 (Right)
A3 or B3	A4 or B4	A3 / B3 (Left) A4 / B4 (Right)	A3 or B4	A4 or B3	A3 / B4 (Left) A4 / B3 (Right)
A3 or C3	A4 or C4	A3 / C3 (Left) A4 / C4 (Right)	A3 or C4	A4 or C3	A3 / C4 (Left) A4 / C3 (Right)
Symmetry Instruction Sets (5 – 8)			**Symmetry Instruction Sets (17 - 20)**		
B1 or A1	B2 or A2	B1 / A1 (Left) B2 / A2 (Right)	B1 or A2	B2 or A1	B1 / A2 (Left) B2 / A1 (Right)
C1 or A1	C2 or A2	C1 / A1 (Left) C2 / A2 (Right)	C1 or A2	C2 or A1	C1 / A2 (Left) C2 / A1 (Right)
B3 or A3	B4 or A4	B3 / A3 (Left) B4 / A4 (Right)	B3 or A4	B4 or A3	B3 / A4 (Left) B4 / A3 (Right)
C3 or A3	C4 or A4	C3 / A3 (Left) C4 / A4 (Right)	C3 or A4	C4 or A3	C3 / A4 (Left) C4 / A3 (Right)
Equivalence Instruction Sets (9 – 12)			**Equivalence Instruction Sets (21 – 24)**		
B1 or C1	B2 or C2	B1 / C1 (Left) B2 / C2 (Right)	B1 or C2	B2 or C1	B1 / C2 (Left) B2 / C1 (Right)
C1 or B1	C2 or B2	C1 / B1 (Left) C2 / B2 (Right)	C1 or B2	C2 or B1	C1 / B2 (Left) C2 / B1 (Right)
B3 or C3	B4 or C4	B3 / C3 (Left) B4 / C4 (Right)	B3 or C4	B4 or C3	B3 / C4 (Left) B4 / C3 (Right)
C3 or B3	C4 or B4	C3 / B3 (Left) C4 / B4 (Right)	C3 or B4	C4 or B3	C3 / B4 (Left) C4 / B3 (Right)

Slow" appeared briefly (cf. Greenwald, Nosek, & Banji, 2003). Subjects were asked at the beginning of the experiment to press the appropriate key on each trial as quickly as possible while trying not to make any errors.

The main body of the IAT, not including practice trials, consisted of 96 tasks. The full list of these is presented in Table 4. For illustrative purposes consider the first row of three cells under the heading "Directly Trained Instruction Sets (1 – 4)." The first cell indicates that the left-side instruction read "If A1 or B1 press left"; the second cell indicates that the right side instruction read "If A2 or B2 press right"; and the third cell lists the four possible target stimuli that could be presented with these two instructions, along with the correct responses (e.g., if the A1 stimulus was presented, pressing the left key was recorded as correct). Each subject was exposed to all 96 trials of the IAT, and both median RTs and ERPs were calculated for each of the three Class–Class conditions and each of the three Class–Nonclass conditions. No significant differences in either RTs or ERPs measures were observed across either of these three sets of conditions, and thus for the purposes of comparison the data were collapsed into just two conditions –a Class–Class condition and a Class–Nonclass condition.

As argued previously, if derived equivalence relations provide a basic model of semantic categories, and the IAT effect is based, at least in part, on the juxtaposition of such categories, RFT would predict that the Class–Nonclass probes should produce longer response times than the Class-Class probes. The results presented in Figure 5 (two left bars) appear to provide support for this prediction, and the statistical analyses did indeed indicate that the response times were significantly greater for the Class-Nonclass than for the Class–Class condition. Interestingly, the grand average ERPs also appeared to be sensitive to the two different IAT conditions, with the Class-Nonclass

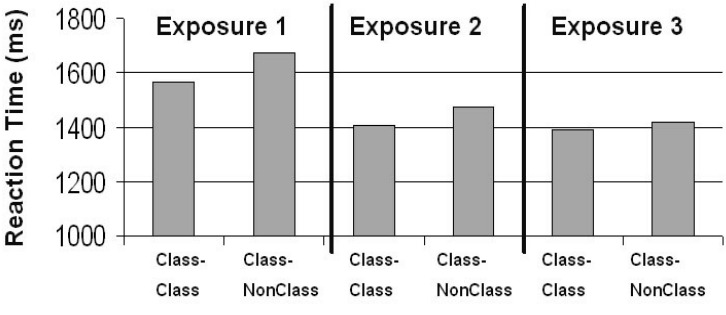

Figure 5. Median reaction times calculated across Class-Class and Class-Nonclass trial-types from the IAT. Results are presented for each of three exposures, with each exposure involving a novel set of stimuli. The difference between trial-types was statistically significant for Exposures 1 and 2, but not for 3.

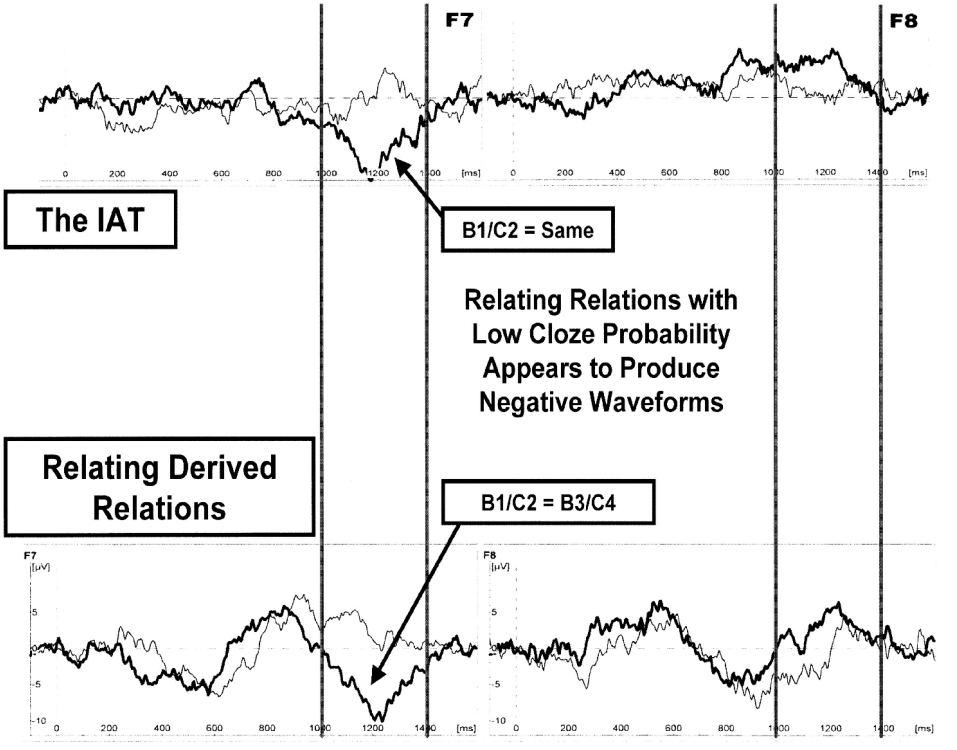

Figure 6. Upper panel: Grand average waveforms calculated across participants for Class-Class (light lines) and Class-Nonclass (dark lines) trial-types from the IAT at electrode sites F7 and F8. Point zero on the graph marks the presentation of the target stimulus on each trial. Significantly greater negative deflections, commencing around 800 ms after target onset, were recorded for the Class-Nonclass trial-types (relative to Class-Class trial-types) at the left frontal site, F7. The opposite pattern was observed at F8, but the difference was non-significant.
Lower panel: Grand average waveforms calculated across participants for relating equivalence relations to equivalence relations (light lines) and relating nonequivalence relations to nonequivalence relations (dark lines) at electrode sites F7 and F8. Point zero on the graph marks the presentation of the comparison stimuli on each trial. Significantly greater negative deflections, commencing around 800 ms after comparison onset, were recorded for the nonequivalence relating trial-types (relative to equivalence relating trial types) at the left frontal site, F7. The opposite pattern was observed at F8, but the difference was non-significant.

probes producing negative waveforms for the left-frontal electrode sites, and positive waveforms for the right-frontal sites. Figure 6 (upper panel) presents the grand averages for electrodes F7 (left) and F8 (right), and in these cases, the waveforms commence around 800 ms following trial onset, and last for about 600 ms. Statistical analyses of

the waveforms across participants indicated that the Class-Nonclass waveforms were significantly more pronounced than the Class-Class waveforms on the left, but this was not true for the waveforms on the right.

Given the absence of any previous ERPs data using natural language categories and the IAT, we should be cautious in interpreting these results, particularly given the preliminary nature of the current research. Assuming, however, that the results are robust, the lateral asymmetry observed for the Class-Nonclass trials on the IAT (i.e., significant negative waveforms on the left but not on the right) may reflect the increased relational or verbal difficulty of these trial-types (see Boroojerdi, Phipps, Kopylev, Wharton, Cohen, & Grafman, 2001; Kolb & Whiteshaw, 2001; Wharton, Grafman, Flitaman, Hansen, Brauner, Marks, & Honda, 2000). To better appreciate this argument, consider one possible RFT interpretation of the two IAT conditions.

On each trial in the Class-Class condition subjects are asked to press a specific key when one of two stimuli in an equivalence relation is presented, and thus it could be argued that the task involves responding to an equivalence relation as equivalent. That is, the task requires that equivalent stimuli be treated as equivalent because they control the *same* key press -this aspect of the procedure is similar to the pREP discussed by Barnes-Holmes, Barnes-Holmes, Smeets, Cullinan, & Leader (this volume). In the Class-Nonclass condition, however, the task involves responding to a non-equivalence relation as equivalent, because now nonequivalent stimuli control the *same* key press. In the Class-Class condition, therefore, one type of relational frame is primarily involved (i.e., equivalence), but in the Class-Nonclass condition two such frames are involved (non-equivalence and equivalence). By definition, therefore, the latter condition involves more relational or verbal responding than the former condition, and thus the increased neural activity observed for the Class-Nonclass condition is entirely consistent with RFT. Although the foregoing interpretation remains highly speculative, it is worth noting that in a completely separate study conducted in the Maynooth RFT laboratory ERPs patterns similar to those produced by the IAT were obtained during a relating relations experiment (Regan, Barnes-Holmes, Steward, Whelan, Barnes-Holmes, Dymond, & Mohr-Pulvermüller, 2004; see Stewart & Barnes-Holmes, this volume, for an extended discussion of relating relations research). In this study, relating nonequivalence relations as equivalent produced greater negative waveforms at left frontal sites than relating equivalence relations as equivalent –the opposite pattern was observed at the right sites, but like the IAT data the differences were statistically nonsignificant (see Figure 6, lower panel). Although these data are preliminary, and any conclusion must remain extremely tentative, this study suggests that the IAT effect may be explained, in part, by the relating of functionally similar versus distinct relational frames. This issue is currently under further investigation by our research group.

Importantly, there are additional RFT explanations for the differential patterns of neural activity observed on the IAT. Specifically, Class-Nonclass trials likely produce relational responses that compete with the correct IAT response. Imagine, for example, a subject who is given the instructions “If B1 or C2 press left” and “If B2 or C1 press right.” As the subject reads the first instruction, B1 may elicit some of the perceptual functions of C1 based on their participation in an equivalence relation. More informally,

seeing B1 makes the subject think of C1 (see Barnes, 1994, for a detailed discussion of equivalence relations and perceptual functions). Thus, the "press-left" function that is instructed for B1 may transfer to C1 via the equivalence relation, which obviously competes with the "press-right" function established for C1 by the second instruction. A similar analysis, in terms of competing relational responses, may be applied to the C2 and B2 stimuli. In RFT terms, therefore, the derived eliciting functions of the stimuli presented on a Class-Nonclass trial may fail to cohere with the two instruction sets. But why should this lack of relational coherence produce increased negative ERP waveforms? One possible explanation is as follows.

Previous research reported in the mainstream neurocognitive literature has shown negative ERPs components to be modulated by the 'cloze probability' (i.e., degree of expectedness) of the final word in a sentence. For example, the sentence, "it is hard to admit when one is *asleep*" elicits a more negative N400 waveform than the sentence, "it is hard to admit when one is *wrong*" (Kutas, 1993; Kutas & Hillyard, 1984). Perhaps the negative waveforms elicited by the Class-Nonclass trials on the IAT indicate that the derived relational incoherence that is produced by these trials is somewhat unexpected relative to the relational coherence produced by the Class-Class trials. Furthermore, this low cloze probability effect could well be enhanced by the "unexpected" requirement to respond to nonequivalence relations as equivalent. Admittedly, the current negative waveform (at F7) occurred later than the N400, but this could be due to the nature of the IAT (e.g., when the target stimulus was presented at time zero, subjects may have re-read one or both instructions before making a response, thus delaying the waveform relative to a task in which the response is to a final word in a single sentence). Insofar as the foregoing interpretation is correct, the current RFT research may have served to highlight a possible functional overlap between the cognitive/verbal activity that occurs during sentence completion tasks, relating relations tasks, and the IAT. In any case, the "low cloze probability" involved in the relating of nonequivalence relations as equivalent, combined with the lack of relational coherence that may occur during Class-Nonclass trials on the IAT, could explain the increased negative waveforms that were observed for these trial-types in the current experiment.

At the present time, it remains to be seen if the patterns of ERP activity recorded using laboratory-induced equivalence relations are also observed using natural language categories. Nevertheless, the preliminary RFT analysis offered here seems quite plausible and may well provide the basis for a behavior-analytic explanation for the IAT effect itself. As an aside, it should be noted that the current RFT interpretation presents a small challenge to the utility of the IAT as an instrument for assessing emotionally valenced social attitudes and the like. On the one hand, previous studies have shown that it is possible to generate equivalence relations by establishing common functions for a number of stimuli (e.g., Smeets & Barnes, 1997), and thus the types of categories typically used with the IAT may be interpreted as pre-experimentally established frames of coordination or hierarchy under contextual control (e.g., *ant* and *abuse* are equivalent in the context of insects and unpleasant things). On the other hand, the current experiment indicates that these frames do not have to contain stimuli with highly emotive functions –arbitrary nonsense stimuli with few functions beyond the relational functions that

define the class are needed (cf. Karpinski & Hilton, 2001). Thus, from a relational frame perspective, the IAT effect may or *may not* reflect specific social attitudes. Of course, differences may yet emerge between arbitrary "nonsense" classes and highly emotive ones using the IAT, but that is another day's work.

Experiment 5. The research described thus far in the current article has demonstrated that both RTs and ERPs appear to provide measures that are sensitive to derived stimulus relations. However, one might well question the need to take ERPs measures at all if they simply reflect RT differences, which was the case in both the semantic priming and IAT studies described above. Indeed, cognitive neuroscientists have also engaged in a similar debate, although they tend to argue from the opposite perspective – why take behavioral measures? (see Wilkinson & Halligan, 2004). Interestingly, it was in conducting the fifth experiment described here that the utility of ERPs measures, even in behavioral psychology, became apparent.

In Experiment 5, the subjects from the previous experiment were provided with an additional two exposures to the equivalence training and testing and IAT. It is important to note that for each exposure a completely new set of nonsense words was employed, and thus any reduction in the relative differences that emerged across exposures could not be attributed to a practice effect with a specific set of stimuli. Rather, those differences would likely reflect improvements in the relational framing activities that are required by the IAT. Indeed, because relational frames are thought to be examples of generalized operant classes, such improvement across exposures should indeed be expected (see Barnes-Holmes & Barnes-Holmes, 2000), as the relevant contextually controlled relational responses become increasingly flexible across multiple exemplars. Interestingly, such improvement was observed to the extent that by the third exposure the response times for both conditions had reduced by approximately 200 ms, and the differences between the Class-Class and Class-Nonclass conditions were no longer significantly different (see Figure 5). In contrast, however, the ERPs data did not show an equally dramatic improvement for the F7 site. Although the location of the negative waveform shifted to the left (i.e., occurring earlier) from the first to third exposure to the IAT for the Class-Nonclass condition, the amplitudes remained significantly greater relative to the Class-Class condition (See Figure 7). In short, the response time and ERPs measures of IAT performance using derived stimulus relations appeared to diverge.

From an RFT perspective the improvement in response time is to be expected (due to the increased flexibility produced by the exemplar training), but this improvement does not necessarily indicate that previously distinct patterns of relational responding have collapsed into a single class, and are no longer functionally distinct. As explained previously, according to RFT the two trial-types on the IAT could involve different patterns of relational responding (i.e., responding to an equivalence relation as equivalent versus responding to an opposite relation as equivalent (see Stewart & Barnes-Holmes, this volume). Insofar as these are functionally distinct patterns of relational responding, and remain so even when both patterns are emitted at similar temporal rates, ideally RFT requires some measure of this difference. The current data, although preliminary and tentative, indicate that ERPs can provide the instrument we need when RT fails us.

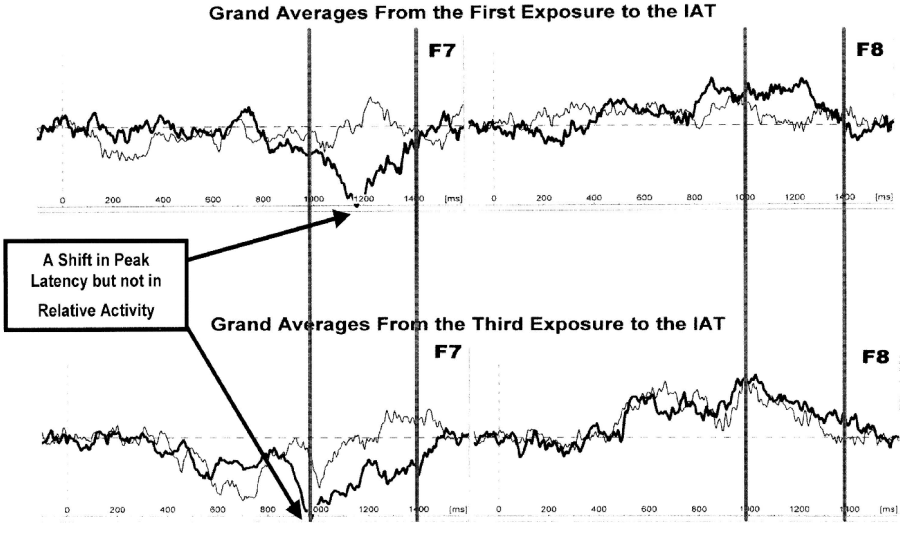

Figure 7. See caption to Figure 6 (upper panel) for details. The upper panel shows the grand average waveforms for the first exposure to the IAT and the lower panel shows these waveforms for the third exposure. Although the peak of the negative waveform for the Class-Nonclass trial types at site F7 shifted towards the left (occurring earlier) during the final exposure, it remained significantly different from the Class-Class trial types. The difference at the right site (F8) was statistically non-significant across all three exposures (second exposure not shown).

Of course, these findings are preliminary, but they do suggest exciting possibilities for RFT research and its interface with cognitive neuroscience. On the one hand, perhaps the differences in neural activity would attenuate, like the response times, with further exposures (there was some evidence of this for the F8 site, although the difference here was not significant during any of the exposures). On the other hand, perhaps the ERPs differences would remain relatively stable. The latter possibility is a great deal more interesting because it would indicate that ERPs could provide a sensitive measure of derived relational responding when response time fails to do so (see Barnes-Holmes, et al., 2001). For this reason alone, RFT research may be well served by adopting, when appropriate, some of the measures and techniques of cognitive neuroscience.

Conclusion

The experimental work described in the first half of the current article replicates the research reported by Hayes and Bisset (1998). Furthermore, it extends that work considerably by demonstrating priming effects with derived stimulus relations using a

single-word lexical decision task, both after (Experiment 1) and before (Experiment 2) a formal MTS equivalence test. The results of Experiment 3 provide additional evidence in favor of a functional overlap between semantic and derived stimulus relations, in that the N400 waveform was shown to be sensitive to the directly trained and equivalent stimulus pairs versus the non-equivalent pairs. Finally, the results of Experiments 4 and 5 indicate that these reaction time and ERPs effects are not restricted to traditional lexical decision tasks, but can also be observed using the IAT. Furthermore, preliminary evidence suggests that ERPs might constitute a more sensitive measure of derived stimulus relations on the IAT than response time.

The current research is important, in that only one published study has undertaken an investigation of the neural correlates of derived relational responding (Dickins, et al., 2001, who used fMRI). Although the experiments described here employed ERPs measures, rather than fMRI, the present findings are broadly consistent with that earlier work in that both measures indicate that derived stimulus relations produce neural effects that are typically observed when humans are engaged in activities that cognitive neuroscientists call semantic processing. Overall, therefore, the findings obtained across all four experiments lend considerable weight to the argument that derived stimulus relations provide a workable behavior-analytic model of semantic relations in natural language.

Future research on derived relations, semantic priming, and implicit associations could employ larger equivalence classes, or more complex relational networks, to more closely model the highly rich and complex semantic relations found in natural language. Indeed, because derived relations are, in a sense, created ab initio in the laboratory, the opportunities for constructing and manipulating networks of stimulus relations, that can then be tested using various priming methods, is almost boundless. Certainly, the results obtained from the current research support the view that the study of derived stimulus relations, combined with some of the procedures and measures of cognitive psychology and cognitive neuroscience, could well provide an important inroad into the experimental analysis of semantic relations in human language.

References

Anderson, J.R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J.R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Balota, D.A. & Lorch, R.F.Jr. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 336-345.

Barnes, D. & Hampson, P.J. (1993). Stimulus equivalence and connectionism: Implications for behavior analysis in cognitive science. *The Psychological Record, 43*, 617-638.

Barnes, D. & Hampson, P.J. (1997). Connectionist models of arbitrarily applicable relational responding: A possible role for the hippocampal system. In J. Donahoe & V.P. Dorsel (Eds.) *Neural-network models of cognition* (pp. 496-521). North-Holland: Elsevier.

Barnes, D. & Holmes, Y. (1991). Radical behaviorism, stimulus equivalence, and human cognition.

The Psychological Record, 41, 19-31.

Barnes, D., McCullagh, P.D., & Keenan, M. (1990). Equivalence class formation in non-hearing impaired children and hearing impaired children. *Analysis of Verbal Behavior, 8*, 19-30.

Barnes-Holmes, D. & Barnes-Holmes, Y. (2000). Explaining complex behavior: Two perspectives on the concept of generalized operant classes? *The Psychological Record, 50*, 251-265.

Barnes-Holmes, D., Barnes-Holmes, Y., Smeets, P.M., Cullinan, V., & Leader, G. (this volume). Relational frame theory and stimulus equivalence: Conceptual and procedural issues. International Journal of Psychology and Psychological Therapy, xx, xxx-xxx.

Barnes-Holmes, D., Hayes, S.C., & Dymond, S. (2001). Self and self-directed rules. In S.C. Hayes, D. Barnes-Holmes, & Roche, B. (Eds.), *Relational Frame Theory: A post-Skinnerian account of human language and cognition* (pp. 119-139). New York: Plenum.

Barsalou, L.W. (1999). Perceptual symbol systems. *Behaviour and Brain Science, 22*, 577-660.

Bentin, S., McCarthy, G., & Wood, C.C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60*, 343-358.

Boroojerdi, B., Phipps, M., Kopylev, L., Wharton, C.M., Cohen, L.G., & Grafman, J. (2001). Enhancing analogic reasoning using rTMS over the left prefrontal cortex. *Neurology, 56*, 526-528.

Branch, M.N. (1994). Stimulus generalization, stimulus equivalence, and response hierarchies. In S.C. Hayes, L.J. Hayes, M. Sato, & K. Ono (Eds.), *Behaviour analysis of language and cognition* (pp. 51-70). Reno, NV: Context Press.

Collins, A.M. & Loftus, E.F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 4007-428.

Cullinan, V.A., Barnes, D., Hampson, P.J., & Lyddy, F. (1994). A transfer of explicitly and non-explicitly trained sequence responses through equivalence relations: An experimental demonstration and connectionist model. *The Psychological Record, 44*, 559-585.

Dasgupta, N., Greenwald, A.G., & Banaji, M.R (2003). The first ontological challenge to the I.A.T: Attitude or mere familiarity? *Psychological Inquiry, 14*, 238-243.

Deacon, T. (1997). *The symbolic species*. Penguin: London.

De Groot, A.M.B. (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behaviour, 22*, 417-436.

De Rose, J.C., De Souza, D.G., Rossito, A.L., & De Rose, T.M. (1998). Stimulus equivalence and generalization in reading after matching to sample by exclusion. In S.C. Hayes & L.J. Hayes (Eds.), *Understanding verbal relations*. Reno, NV: The Context Press.

Devany, J.M., Hayes, S.C., & Nelson, R.O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior, 46*, 243-257.

Dickins, D.W., Singh, K.D., Roberts, N., Burns, P., Downes, J., Jimmieson, P., & Bentall, R.P. (2001). An fMRI study of stimulus equivalence. *Neuroreport, 12*, 2-7.

DiFore, A., Dube, W.V., Oross III, S. Wilkinson, K., Deutsch, C.K., & Mcilvane, W.J. (2000). Studies of brain activity correlates of behavior in individuals with and without developmental disabilities. *Experimental Analysis of Human Behavior Bulletin, 18*, 33-35.

Dugdale, N. & Lowe, C.F. (2000). Testing for symmetry in the conditional discrimination of language trained chimpanzees. *Journal of Experimental Analysis of Behavior, 73*, 5-22.

Fields, L. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior, 48*, 317-332.

García, A. & Benjumea, S. (2001). Pre-requisitos ontogenéticos para la emergencia de relaciones si-

métricas. *International Journal of Psychology and Psychological Therapy, 1,* 115-135.

Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464-1480.

Greenwald, A.G., Nosek, B.A., & Banaji, M.R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.

Hayes, S.C., Barnes-Holmes, D., & Roche, B. (2001). *Relational Frame Theory: A post-Skinnerian account of human language and cognition*. New York: Plenum.

Hayes, S.C., & Bisset, R.T. (1998). Derived stimulus relations produce mediated and episodic priming. *The Psychological Record, 48*, 617-630.

Hayes, S.C., & Hayes, L.J. (1992). Verbal relations and the evaluation of behavioural analysis. *American Psychologist, 47*, 1383-1395.

Hill, R., Strube, M., Roesch-Ely, D., & Weisbrod, M. (2002). Automatic vs. controlled processing in semantic priming-differentiation by event-related potentials. *International Journal of Psychophysiology, 44*, 197-218.

Holcomb, P.J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language, 35*, 66-85.

Holcomb, P.J. & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related potentials. *Language and Cognitive Processes, 8*, 379-411.

Holcomb, P.J. & Neville, H.J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology, 19*, 286-300.

Horne, P.J. & Lowe, C.F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65*, 185-241, 341-353.

Kolb, B., & Whiteshaw, I.Q. (2001). *An introduction to brain and behaviour*. NY: Worth Publishers.

Karpinski, A. & Hilton, J.L. (2004). Attitudes and the implicit association test. *Journal of Personality and Social Psychology, 81*, 774-788.

Kounios, S.A. & Holcomb, P.J. (1992). Structure and process in semantic memory: evidence from event-related potentials and reaction times. *Journal of Experimental Psychology General, 121*, 460-480.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for simple-word and sentence context effects. *Language and Cognitive Processes, 8*, 533-578.

Kutas, M. & Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 1161-1163.

Leslie, J.C., & Blackman, D.E. (2000). *Experimental and applied analysis of human behavior*. Reno, NV: Context Press.

Lipkens, R., Hayes, S.C., & Hayes, L.J. (1993). Longitudinal study of the development of derived relations in an infant. *Journal of Experimental Child Psychology, 56*, 201-239.

Luciano, M.C., Barnes-Holmes, Y., & Barnes-Holmes, D. (2001). Early development history and equivalence relations. *International Journal of Psychology and Psychological Therapy, 1,* 137-149.

McClelland, J.L. & Rumelhart, D.E. (1988). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*. Cambridge, MA: MIT Press.

McNamara, T.P. & Altarriba, J. (1988). Depth of spreading activation revisited: semantic mediated

priming occurs in lexical decisions. *Journal of Memory and Language, 27*, 545-559.

Meyer, D.E. & Schvaneveldt, R.W. (1971). Facilitation in recognition pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227-234.

Neely, J.H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G.W. Humphreys (Eds.), *Basic processing in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Lawrence Erlbaum Associates.

Neely, J.H., Keefe, D.F., & Ross, F. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectations and retrospection semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1003-1019.

Phelps, E.A., O'Connor, K.J., Cunningham, W.A., Funayama, E.S., Gatenby, J.C., Gore, J.C., & Banaji, M.R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12*, 729-738.

Reese, H.W. (1991). Mentalistic approaches to verbal behavior. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (pp. 151-177). Reno, NV: Context Press.

Regan, D., Barnes-Holmes, D., Stewart, I., Whelan, R., Barnes-Holmes, Y., Dymond, S., & Mohr-Pulvermüller, B. (2004). *Testing the RFT model of analogical reasoning: Reaction times and event related potentials*. Paper presented at the Annual Conference of the Experimental Analysis of Behaviour Group, London, England.

Shusterman, R.J. & Kastak, D. (1993). A california sea lion (Zalophus californianus) is capable of forming equivalence relations. *The Psychological Record, 43*, 823-839.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research, 14*, 5-13.

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M.D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213-245). Hillsdale, NJ: Erlbaum.

Sidman, M. (1994). *Stimulus Equivalence: A Research Story*. Boston: Authors Cooperative.

Smeets, P.M., & Barnes, D. (1997). Emergent conditional discrimination in children and adults: Stimulus equivalence derived from simple discriminations. *Journal of Experimental Child Psychology, 66*, 64-84.

Stewart, I. & Barnes-Holmes, D. (this volume). Relational frame theory and analogical reasoning: Empirical investigations. *International Journal of Psychology and Psychological Therapy, xx*, xxx-xxx.

Weisbrod, M., Kiefer, M., Winkler, S., Maier, S., Hill, R., Roesch-Ely, D., & Spitzer, M. (1999). Electrophysiological correlates of direct versus indirect semantic priming in normal volunteers. *Cognitive Brain Research, 8*, 289-298.

Wharton, C.M., Grafman, J., Flitaman, S.S., Hansen, E.K., Brauner, J., Marks, A., & Honda, M. (2000). Towards neuroanatomical models of analogy: A positron emission tomography study of analogical mapping. *Cognitive Psychology, 40*, 173-197.

Wilkinson, D. & Halligan, P. (2004). The relevance of behavioural measures for functional-imaging studies of cognition. *Nature Reviews: Neuroscience, 5*, 67-73.

Received January 15, 2004
Final acceptance June 10, 2004