ON AIMS AND METHODS IN THE NEUROIMAGING OF DERIVED RELATIONS

DAVID W. DICKINS

UNIVERSITY OF LIVERPOOL

Ingenious and seemingly powerful technologies have been developed recently that enable the visualization in some detail of events in the brain concomitant upon the ongoing behavioral performance of a human participant. Measurement of such brain events offers at the very least a new set of dependent variables in relation to which the independent variables familiarly manipulated in the operant laboratory may be explored. Two related paradigms in which a start has been made in such research concern the derivation of novel or emergent relations from a baseline set of trained relations, and include the phenomenon of transitive inference (TI), observed in studies of stimulus equivalence (SE) and serial learning (SL) or seriation. This paper reviews some published and forthcoming neuroimaging studies of these and related phenomena, and considers how this line of research both demands and represents a welcome synthesis between types of question and levels of explanation in behavioral science that often have been seen as antithetical.

Key words: stimulus equivalence, transitive inference, functional magnetic resonance imaging, derived relations, matching to sample, serial learning, brain regions

FUNCTIONAL MEASURES OF
BRAIN EVENTS

Since the mid-1980s it has been possible, using positron emission tomography (PET) or functional magnetic resonance imaging (fMRI), to obtain average neuroimaging signals during blocks, tens of seconds in duration, within which there occur behavioral trials of specific kinds. The validity of such methods depends heavily upon a plausible rationale for making comparisons between related types of behavioral trial, and is subject to the limitations of spatial smoothing of data and deriving results from groups of participants on the assumption that their brain anatomy, having been spatially aligned, can be meaningfully compared. This paper reviews some of these block-designed neuroimaging studies, which Postle and D'Esposito (2000) refer to as "first generation," with the aim of informing future experiments using the "second generation" event-related fMRI. In practice, the greater power of certain kinds of comparison made possible by block-designed studies still can be used to elucidate the findings from event-related investigations, as we hope to demonstrate with the complementary pair of forth-

coming experiments from our laboratories, described in the final section.

In one of a series of related articles published in the *Experimental Analysis of Human Behavior Bulletin*, McIlvane and his colleagues (Deutsch, Oross, DiFiore, & McIlvane, 2000) give the following quotation from Skinner (1989).

There are two unavoidable gaps in any behavioral account: one between the stimulating action of the environment and the response of the organism and one between consequences and the resulting change in behavior. Only brain science can fill those gaps. In doing so, it completes the account; it does not give a different account of the same thing. (p. 18)

The articles encourage the community of behavior analysts to follow the lead of the cognitivists and avail themselves of new developments in brain imaging. An interesting talk given by Bill McIlvane in the U.K. further stiffened my resolve to beg, successfully, for some space on the scanner at the Walton Centre for Neurology and Neurosurgery and initiated my own involvement with this technology.

Taking the above quotation in isolation, my first impulse was to declare some disagreement with Skinner, and to assert that I preferred the views of Niko Tinbergen. In a classic paper, Tinbergen (1963) argued that there are four complementary levels of explanation, all of which need to be in place for a full un-

Correspondence concerning this paper should be addressed to David W. Dickins, School of Psychology, University of Liverpool, Eleanor Rathbone Building, Bedford Street South, Liverpool L69 7ZA, United Kingdom (e-mail: dickins@liverpool.ac.uk).

doi: 10.1901/jeab.2005.92-04

derstanding of behavior. However, immediately after the two sentences quoted above, Skinner goes on to say:

Human behavior will eventually be explained (as it can only be explained) by the cooperative action of ethology, brain science, and behavior analysis. The analysis of behavior need not wait until brain scientists have done their part. The behavioral facts will not be changed, and they suffice for both a science and a technology. Brain scientists may discover other kinds of variables affecting behavior, but they will turn to a behavioral analysis for the account of the effects of these variables. (Skinner, 1989, p. 18)

The fact that two of Tinbergen's levels of explanation are evolutionary would have been entirely congenial to Skinner in his latter days, during which he became enthusiastic about cross-connections and parallels between his own ideas and those of Darwinian theory. In fact Tinbergen (1963) laid down what are still regarded as the definitive set of four *why* questions, or levels of explanation, in ethology (behavioral biology), making the point that satisfactory answers to all four of them were needed to give a complete account of what gave rise to and what maintains a given pattern of behavior found in an organism. Two of these levels of explanation were at the proximate level, applying to the particular organism currently under study: (1) *causal*—the explanation of how the machinery of the animal's brain, muscles etc. combined their activities to generate instances of the behavior, and (2) *ontogenetic*—the consideration of the individual developmental history of the organism, both in terms of genetic programming and how it interacted with its environment. The other two *whys* were at the ultimate or evolutionary level, namely, (3) *phylogenetic*—asking what was the evolutionary history of the behavior, through what conceivable ancestral stages might it have evolved (a major concern of comparative cognition); and (4) *functional*—how does the behavior (if it does) help in adapting the organism to its natural environment, namely, enhancing its chances of survival and/or opportunities to reproduce, its reproductive success being the ultimate measure of its Darwinian fitness.

Prior to any kind of explanation it is of course essential to obtain an accurate description of the behavior(s) to be explained.

To the ethologists it was axiomatic that this was based upon observation of animals freely behaving in their natural habitat, or environment of evolutionary adaptedness (Bowlby, 1969, 1973), but they recognized the value of experimental manipulation of features drawn from knowledge of this context, as in their analyses of the key features of salient stimuli to which animals may maximally respond. It would be interesting to attempt a study of the deployment of stimulus equivalence (SE) in ordinary life, but the discrimination between previously reinforced and novel emergent behaviors would be hard to document.¹

Neuroimaging studies are clearly aimed at explanation type 1 (causal), but where do reinforcement theory and cognitive explanations fit into this schema, and are they mutually exclusive, or are they complementary? The behavior-analytic approach is surely mainly a type 2 developmental history of how behavior comes about, with emphasis upon individual learning via reinforcement contingencies, whereas cognitive psychology is primarily concerned with type 1, though, like behavior analysis, it often tends to remain somewhat parochially within its own seemingly distinct universe of discourse.

Cognitivists often stress the key role of representations, which sound like ghostly images or consciously available recollections of stimuli and perhaps planned actions, but the term is simply a reference to the fact that it is (some of) the information contained in the stimuli and in the responses that is the common currency of the brain's activities. Externally, cognitive psychologists contrive artificial stimulus situations and measure aspects of the response in an attempt to chart and codify the natural history of cognition. The combination of the stimulus situation and the requisite responses is commonly referred to as a task, and in order to be able to generate appropriate responses the participants may be given instructions, incorporating rules, or they may be required to arrive at such rules themselves, either implicitly, without being able to say what these rules are if asked, or explicitly, where they are able correctly to

¹ Can the reader remember realising that the term "cuckold" is derived etymologically from the onomatopoeic bird-name "cuckoo" because of the common feature of delegation of raising the young? Is this a semantic example of transitivity?

specify these rules. SE and serial learning (SL) seem allied to phenomena such as inference, which to a cognitive psychologist entails the computation of remembered information, following a set of rules, by means of which a further outcome is generated. No clear divide necessarily exists between implicit and explicit problem solving, since the distinction may simply rest upon the efficacy of communication between participant and experimenter, or in some sense within the participant, but as evaluated by the experimenter.

If cognitivists obtain consistent results they try to map out, often as an information flow or control-systems engineering diagram with boxes and arrows, a series of minimal operational steps by means of which the organism is able to retain and compute the information in the stimuli in order to organize an appropriate response. Although these steps may start with high-level, virtually folk-psychological (Stich, 1983) or mentalistic natural language descriptions, they may lead to the generation of computational models that have the status of theory because, in principle, they can be tested in various ways. Individuals may exist who seem to have difficulties with one limited part of the computational system. They may have developmental abnormalities, or be the victims of brain damage of one sort or another, the latter providing most of the material for classic cognitive neuropsychology. It may be possible empirically to block parts of the system by competitive interference (by means of simultaneous tasks), or with transcranial stimulation (Jahanshahi & Rothwell, 2000; Nixon, Lazarova, Hodinott-Hill, Gough, & Passingham, 2004: but heed warnings about TMS in Walsh & Pascual-Leone, 2003). Computer simulations of the model can be conducted which may generate specific features of actual behavior, or be modified until they do. Any plausible account, fitting all of the features just described, may hold the scientific stage for a while as the best available model of how the given task is performed, with the proviso that a cat can often be skinned in a variety of ways.

If there are multiple plausible models from which to choose, scientists prefer the simplest. But the nub of neuropsychology is that the sequence of operations posited for the solution of a particular task should veridically (and verifiably) map on to actual processes in the

brain. It may not be the case that an organ with such a complex evolutionary history itself conforms to some principle of parsimony. As Donald Hebb once reminded us:

Because a simple task could, theoretically, be handled by a simple mechanism does not mean in fact that the brain handles it that way. In an uncomplicated nervous system, yes; but in the complex brain of a higher animal other mechanisms may insist on getting into the act and turn the simple task into a complex one. (Hebb, 1958, p. 453).

In short, our aim is to use imaging to attempt to track how the brain “does” derived relations. But what exactly does fMRI measure?

Ironically, we must await further developments in understanding what fMRI does and can measure. We know that directly it logs changes in the proportion of oxyhemoglobin and deoxyhemoglobin molecules in the blood, and that the BOLD (blood oxygen level derived) response probably corresponds to an influx of more oxygenated blood as a response to some local mechanism signaling need. That this need reflects enhanced activity of neurons presumably relevant to their information-handling actions has been given greater credence by the elegant correlations between the BOLD response and microelectrode recordings in response to a checkerboard pattern demonstrated at the same time in the same part of an anaesthetized monkey's brain by Logothetis, Pauls, Augath, Trinath, and Oeltermann (2001). But this is not the whole story. There is a huge problem here of scale. As R. S. Menon (2001) points out, in a single 1 mm^3 voxel, the smallest volume of brain which the most sensitive fMRI machine can report upon individually, one might typically find 10^5 neurons and 10^8 synapses. Also it is not always easy to discriminate between oxyhemoglobin determinations from relatively localized vessels and those from larger vessels drawing blood into larger regions. Even the typical shape of the BOLD response itself, which transpires over several seconds, and may vary considerably from region to region of the brain, may sometimes but not always show an initial negative dip in activity before the characteristic rise, the significance of which is in dispute. In due course, results from today's neuroimaging investigations will enable the construction of better models more capable of

predicting patterns of activity across the brain and integrating the piecemeal acquisition of understanding based upon the methods of subtraction discussed in this paper. Such models will doubtless employ new methods of measuring brain activity, to complement or transcend the present indirect evidence supplied by fMRI.

With these caveats in mind we will review the starts that have been made in imaging derived relations, and sketch some experiments under construction. It is best to start, however, by characterizing the behaviors that are being investigated, and defining in fairly strict operational terms the behavioral laboratory methods and procedures involved, and the results obtained, and then moving to various views and explanations of these outcomes that have been proffered in the literature. We will extract from this account some degree of rapprochement between behavior-analytic and cognitivist approaches, which is necessary in order to formulate meaningful questions about the role of the brain in the organization of this behavior.

DERIVED RELATIONS

Groups of stimulus-stimulus relations trained between arbitrary pairs of stimuli by operant techniques may form the basis from which new, untrained relations between stimuli from different pairs may be derived, provided that the two stimuli in the emergent pair are linked indirectly to each other via one or more of the other previously trained linkages. Two such paradigms, that of stimulus equivalence (SE) and serial learning (SL) or seriation, will be described.

In simple terms, stimulus equivalence (SE) is based on learning arbitrary connections between stimuli in specific pairs, and then being able to construct new pairings from these without specific instruction. Thus if a participant were taught that *orange* went with *square*, and *square* went with *violin*, then, without further training or instruction the subject could reverse these pairings and be able to say for example that *square* goes with *orange* and *violin* goes with *square*. This capacity to reverse the direction of the trained relations is labelled *symmetry*. The subject could also make a *transitive* leap by saying that *orange* goes with *violin*. And one can combine symmetry

and transitivity by saying that *violin* goes with *orange*. If all these new connections could be derived from the two trained relations, then the three stimuli can be said to be members of a common equivalence class or set. In a number of ways the stimuli within such a set have become interchangeable, which might be formally represented as follows:

$$\text{ORANGE} \equiv \text{SQUARE} \equiv \text{VIOLIN}$$

In practice, there would have to be at least one other separate similar set so that choices between alternative stimuli could be arranged.

Again, in simple terms, in SL some specific relations between stimuli first have to be learned by encountering them on a number of occasions in specific pairs. In each pair a particular stimulus must be selected. The pairs overlap in a particular way, so that a particular stimulus may enter into two different pairs, being the correct one to select in one pair, but incorrect in the other. When the stimuli are then paired in entirely new combinations, new relations between them can be derived from the initial learning. Given *orange* and *square* a number of times, participants might learn always to choose *square*, while at the same time given trials with *square* and *violin*, they might learn to choose *violin*. When they have learned these, the only relation they could derive here, in the sense of a consistent choice to make without further instruction, would be that when *violin* is now paired with *orange*, one should systematically choose *violin*. However, this could be due to discriminating between *orange*, the choice of which was never reinforced in training, and *violin*, which always was reinforced. A simple correction to circumvent this elementary confound is to deploy a longer series of at least five stimuli making four overlapping pairs, such as:

$$\begin{aligned} &\text{ORANGE} - \text{SQUARE}, \text{SQUARE} - \text{VIOLIN}, \\ &\text{VIOLIN} - \text{CRAB}, \text{CRAB} - \text{NECKTIE} \end{aligned}$$

Here the crucial demonstration of transitive inference (TI) would be to present *square-crab*, from which *crab* should definitively be selected. What has happened here is that from choices

learned in restricted (and randomly presented) series of training pairs, the participant has in effect imposed a serial order or hierarchy on these stimuli which might be formally represented as:

ORANGE < SQUARE < VIOLIN
< CRAB < NECKTIE

It is difficult to give clear-cut, real-world examples of such derived relations. SE and SL are formal laboratory-controlled processes intended to help us illuminate scientifically the underlying structure of the complexities of everyday life. In practice, these are likely to involve many kinds of relations, some derived and some based on a history of direct reinforcement. SE, however, does seem to map on to words and their referents (T. E. Dickins & Dickins, 2001; Place, 1995) in the sense, for example, that a set of symbols—such as the word *dog* spoken and heard, written and read, and corresponding words in other languages—are in equivalence relations with each other, and in addition they become attached to a variety of signs for dogs, that is to say stimuli associated with dogs, such as the sight, sounds, and smells of dogs. The symbols and signs all together constitute an equivalence class. Note that the relation between the symbols and the canine signs is arbitrary, as is true for most words and their referents. Children having learned to say *dog* in the presence of a dog will spontaneously be able to point to a picture of a dog when they hear *dog* spoken (involving symmetry). If they also learn to read the written word *dog* (i.e., when they see this combination of letters they respond with the spoken word *dog*), then they will be able in an appropriate future context to indicate directly a dog or a picture of a dog on seeing the written word (transitivity).

SL that involves the process of comparisons on a scale obviously would apply to a child learning symbols for measures, for example of length, and learning that miles are much bigger than inches, since mile > yard > foot > inch. SL could apply also to social learning, as in speculations about the adaptive functions of SL in other species. Thus a child might learn degrees of authority between more dominant and more submissive peers, and

hierarchies of adult figures such as teachers and head teachers in her social world. We have to consider cases where her relational properties of stimulus objects are arbitrary and therefore to some extent must be experience based and not simply apparent, as in size or length judgements.

STIMULUS EQUIVALENCE (SE)

In the SE paradigm the trained relation between the stimuli in a pair typically is based upon some kind of conditional discrimination, though derived relations can sometimes be obtained when the two stimuli in each pair are simply presented successively in Pavlovian (or sensory preconditioning) fashion a number of times, without reinforcement consequences (Leader, Barnes-Holmes, & Smeets, 2000). In a conditional discrimination, whether or not a particular behavior is reinforced in the presence of one stimulus depends upon whether or not a second stimulus is also present. A common form of conditional discrimination usually adopted in studies of SE is the so-called² matching-to-sample (MTS) procedure. Here, in the presence of, or following³

²The appellation “matching” derives from the procedure of identity matching (ID-MTS) in which a sample, such as X, is followed by presentation of a choice between X and Y, where X is the correct choice and is an identical reiteration of the sample stimulus. In SE, training is done with purely arbitrary pairings of stimuli, not derivable from any pre-experimental experiences associating the stimuli in question. Confusingly in SE, identity matching is said to occur spontaneously for the stimuli involved after they have been linked in an equivalence class. Technically this is one of the three tests that should be applied to satisfy the three properties of an equivalence set in mathematical set theory, but this is seldom done in practice. This is largely because identity matching is such a simple and familiar task for humans. Starting from scratch, ID-MTS is much simpler for humans than arbitrary MTS, because at most, only short-term or working memory is required (in the zero delay procedure) in identity matching, whereas sample-comparison pairings need to be committed to long-term memory in arbitrary MTS. Interestingly, for pigeons these two tasks can be mastered with equal facility, but whereas humans can generalize from training with a few exemplars on ID-MTS to new stimuli, whether or not pigeons show generalized identity matching depends upon crucial details of training procedures (A. Wright, 2001).

³In all of the SE studies described in this review, a delay procedure is used in which sample presentation precedes presentation of the comparison(s). Typically, the offset of sample coincides with the onset of the comparison(s), a zero delay, but in some cases described below a non-zero delay was employed.

presentation of (say) stimulus A1, a simultaneous discrimination between (say) B1 and B2 is set up. A response to B1 is reinforced, but if B2 is selected then reinforcement is withheld, and/or differential feedback is given, with choices of B1 being labeled correct but choices of B2 being labeled incorrect. The first stimulus, such as A1, is usually termed the sample, with B1 being the correct comparison stimulus, whereas B2 is an incorrect comparison or foil. In parallel with such sample-comparison training, which can be symbolized as A1-B1 training, a parallel instance of A2-B2 training also will be given, so that the participant learns to select B2 rather than B1, following a presentation of A2.

An alternative method is the stimulus pairing yes/no procedure (Fields, Reeve, Varelas, Rosen, & Belanich, 1997) or the Relational Evaluation procedure (REP) (Cullinan, Barnes-Holmes, & Smeets, 2001). Here, presentation of A1 is followed by one comparison only, but two responses are available, such as the choice between one key or button representing *yes* (or *same*) and the other representing *no* (or *different*)⁴. Feedback is given such that if, after presentation of A1, B1 is presented, the *same* response is deemed correct, whereas if, after presentation of A1, B2 is presented, the *different* response is deemed correct (and vice versa after presentation of A2 as the sample).

Following any training method, a number of such relations are established, the simplest case of which would be A1-B1, A2-B2, and B1-C1, B2-C2. Note that two stimuli, B1 and B2, have connections with two other stimuli, which defines them as *nodes*. Potentially this training will lead to the formation of two SE classes (both defined and identified as explained below), namely, A1B1C1 and A2B2C2. Experiments may entail more than two nodes per potential SE class, such as an ABCDE 3-nodal design, and may of course involve more than two such training classes.

After trained relations have been thoroughly established, tests are conducted in which any

reinforcement or informational outcome is withheld after the requisite choice. These tests may be used to probe for emergent new relations between stimuli. Such novel tests comprise symmetry (reversals of the trained sample-comparison pairings, such as B-A and C-B trials); transitivity (such as A-C trials); and the combination of symmetry and transitivity, or *equivalence* trials (such as C-A). Symmetric and transitive relations between members of a group of stimuli such as A, B, and C, map on to the mathematical relation in set theory of an equivalence set (Sidman & Tailby, 1982). Logically this should also entail passing tests of identity (reciprocity), such as A-A, B-B, or C-C, though this is seldom tested in practice. (The use of identity matching as a control test in fMRI studies is discussed below.)

In practice, tests of symmetry are usually passed from their inception, but participants may take several (unreinforced) trials before they consistently choose the comparison corresponding to transitive or equivalence relations, and it is common to find that the *yield* of an SE experiment (i.e., the proportion of a sample of participants consistently passing tests involving transitivity before the end of the experiment) is less than 100%. Those who do pass, however, thereby demonstrate the emergence of many new relations from a limited set of trained relations. This ratio of new to old shows an exponential gain as a function of the initial number of trained relations linked by nodes: from an initial two relations (AB, BC), four new relations, excluding identity, emerge; from three (AB, BC, CD), nine emerge – three symmetry, three transitivity, and three equivalence; from four (AB, BC, CD, DE) 16, etc. Passing these tests, as described, is easy to perceive as the consequence of a new structural interrelation of stimuli, especially since stimuli within a particular SE class display other signs of relatedness, such as being able now to substitute for each other as discriminative stimuli for the control of subsequently acquired behaviors: if, say, A1 is established as a discriminative stimulus, then without further training B1 and C1 also will function as discriminative stimuli for the same behavior. The stimuli here are showing *functional equivalence*. Other common properties accruing to members of an SE class are described below.

⁴Cullinan et al. (2001) found the terms “same” and “different” to be crucial for good yield, but Fields (personal communication) compared these with Yes/No and found no significant difference in yield. Much may depend on other aspects of the instructions and other expectations (see further discussion in text).

DERIVED RELATIONS: SL

Another paradigm in which many new relations can be derived from a limited set of previously trained relations is that of serial learning (SL) or seriation. Here each pair of stimuli is presented as a simultaneous discrimination. These are also conditional in the sense that pairing determines which stimulus is the correct one to select. Thus if A is paired with B, B may be the correct choice, but if C is paired with B, C will be the correct choice. A linear series of such choices is set up, but the individual pairs for successive trials may be presented either in an ordered or in a random sequence. For the sake of exposition we may characterize the polarization of choices using $<$, the *less than* sign, with the stimulus to the right of this symbol always being the correct choice in each pair. The linear series would be represented, in the simplest case (see below) as: $A < B$, $B < C$, $C < D$, $D < E$. Again, once participants have mastered these individual conditional discriminations, often presented in random order until a criterion of consistent correct responding is attained, they then are given novel tests (again meaning non-reinforced trials) of non-adjacent pairs. Such non-adjacent pairs provide tests of transitivity, in the SL sense of this term, but in the above small set only one such test, BD, is an indisputably valid test of transitivity.

All other available non-adjacent pairs will contain either an E or an A stimulus (or both). Now, in training, E will always have been the correct choice, and A will never have been correct, and on this basis selection of E in any pair, such as CE, or of the other, *not-A* stimulus in any pair such as AC, can be predicted without assuming relations which embrace the stimulus set. A longer series, such as $A < B$, $B < C$, $C < D$, $D < E$, $E < F$, $F < G$, $G < H$, $H < J$, $J < K$ would enable transitivity tests of non-adjacent stimuli, such as EG, separated by one stimulus in the sequence, or CF, separated by two, or BG by three, etc., without deploying either end stimulus (A or K). Acuna, Sanes, and Donoghue (2002) ran such tests using a series of 11 shapes. After mastery of the adjacent training pairs that had been presented in random order, their participants were able to pass these tests (in the absence of

any formal reinforcement or specific guidance) as if, in cognitive⁵ terms, they had constructed some reference list or mental line, which organized all of the stimuli into an ordinal series.

COMPARISON OF PARADIGMS:
SE VERSUS SL

These two paradigms have been selected for this paper because (a) they show intriguing empirical similarities and differences from one another; (b) one of them, seriation, has been the subject of several attempted levels of explanation, both in terms of reinforcement theory (e.g., Wynne, 1995) and in terms of cognitive mechanisms (e.g., Acuna, Sanes, & Donoghue, 2002). Such levels of explanation may serve as exemplars (and/or caveats) for a similar analysis which needs to be attempted with SE, in order to formulate and attack appropriate questions about how the brain “does” SE; and (c) both have been the subject of neuroimaging studies (see below).

Logical similarities and differences. The empirical definition of SE typically follows that of Sidman and Tailby (1982) and adopts the three criteria of reflexivity, symmetry, and transitivity, that appear to operationalize the defining properties of an equivalence set in mathematical set theory. A parallel set of logical properties for the ordered relations of SL (B. C. Wright, 2001, citing Stevens, 1951) would include transitivity but specifically exclude reflexivity and symmetry. A fourth property, of connectedness, is that the relation in question applies to all the members of the set, which is true of both SE and SL.

Phylogenetic differences. Transitive responses following SL training can be shown in many non-human species (e.g. monkeys: McGonigle & Chalmers, 1992; Treichler, Raghanti, & Van Tilburg, 2003; corvids: Lazareva et al., 2004; Paz-Y-Mino, Bond, Kamil, & Balda, 2004, and many other studies cited in Wynne, 1995), whereas with the notable exception of some studies of a California sea lion⁶ (Kastak, Schusterman, & Kastak, 2001; Schusterman &

⁵Pejorative connotation not intended – see ensuing discussion.

⁶The sea lion was trained to make some symmetric and transitive responses by giving reinforcement for correct responses on such tests before making correct responses spontaneously on the first test trial (before any reinforcement had been signalled) of similar tests of entirely novel relations.

Kastak, 1993, 1998) there exists no formal evidence for SE formation, defined as the concurrent spontaneous occurrence of symmetry, transitivity, and identity, in any species other than humans, even if some performances of chimpanzees may be considered tantamount to the formation of SE classes (Cerutti & Rumbaugh, 1993).

Functional speculations. We should be careful, though, not to think simply in terms of human/non-human dichotomies unless we are seriously concerned with understanding species-specific qualities in our own species. Questions of capacity (in some general sense) are less evolutionarily apt than more specific questions of utility. For SL, one may conjecture that there are many circumstances in which social animals might gain by ranking others in the group in dominance terms (Pazy-Mino et al., 2004), or foragers might benefit by ranking food items in some serial order of preference. For SE, Schusterman and Kastak (1998) have suggested that SE might function to enhance and sustain mother–infant bonds, which are highly developed in the sea lion, by facilitating the integration of individual stimuli associated with each other—such as, for the pup, the smell, voice, and appearance of its mother, together with the reinforcing stimuli associated with suckling and warmth that she provides. Similarly, the recognition of lineages, kin groups, and alliances might be assisted by SE in highly social species. Perhaps the higher neocortical index (amount of neocortex in relation to the overall size of the brain) as a function of increasing group size found in several mammalian orders, including primates (Dunbar, 1992, 1993), confers cognitive options for the formation of derived relations and may thus be a preadaptation or exaption for the evolution of language (T. E. Dickins & Dickins, 2002).

Opposite effects of nodal distance in SE and SL transitivity. As noted above, the emergence of transitive relations in SE may not be immediate. Such a ‘delayed emergence’ (Sidman, 1994) is objectively indicated by (i) the higher proportion or probability of incorrect responses on early trials prior to the stabilization of correct responses on later trials⁷, and (ii) the finding that response latencies in transi-

tivity and equivalence tests are typically longer at first than those on tests of trained relations or symmetry (e.g., Bentall, Dickins, & Fox, 1993). There is also a ‘nodal distance’ (Fields & Verhave, 1987) or ‘nodal number’ effect⁸: response latencies in transitivity and equivalence increase as a linear function of the number of nodes between sample and comparison, at least over the first few trials (Bentall, Jones, & Dickins, 1999) or over a longer period (Spencer & Chase, 1996). Imam (2001) has queried these findings and the indication they appear to give of an SE structure in relation to which the implied retrieval of derived relations may take place. His use of just two or three individual participants rather than groups, and the complication of a limited hold requiring a tradeoff between speed and latency, however, weakens rather than strengthens support for his rejection of this kind of account.

Exactly the converse effect of the number of intervening stimuli is commonly found in SL (Wynne, 1995). Described as the “symbolic distance effect,” the response latencies become shorter, and the accuracies higher, the more intervening stimuli in the series there are between the nonadjacent test pair. In SL there is also a ‘serial position effect’: when response accuracies of training pairs are plotted against their position in the order of the implicit series, a U-shaped relation is observed, with the more accurate responding at the ends and the least accurate in the middle, even though pairs are presented in random sequence during training (Wynne, 1995).

THEORETICAL ACCOUNTS OF SL AND SE

We have characterized the trained relations from which other kinds of relations may be derived as conditional discriminations. A conditional discrimination is defined in terms of the

stable incorrect responses after repeated tests of transitivity. The reaction times of these responses also showed a systematic decline with repeated tests as is often the case during the stabilization of correct responses.

⁸Imam (2001) prefers nodal number to nodal distance. The latter term does mesh with the idea of some extended, quasi-spatial structure, a cognitive representation perhaps, but “number” too could be regarded as an overinterpretive metaphor expressing the same concept verbally. All language is metaphor.

⁷It is not invariably correct responses that stabilize. Holth and Arntzen (2000) report participants who showed

contingencies that apply, of reinforcement or informative consequences, for a particular response to one particular stimulus (a comparison) dependent upon which other stimulus (the sample) is present or has just been presented. So far we have been neutral about how the participant acquires these trained relations in the context of earlier learning and parallel learning.

The fact that similar phenomena are found in SL in human and nonhuman subjects has encouraged the search for a parsimonious explanation of seriation. Wynne (1995) has shown how far general reinforcement theories such as those of Bush and Mosteller (1955) and Rescorla and Wagner (1972) can account for these phenomena. The former explained quite a bit, and the latter still more, but neither can account for the following finding, without making further special assumptions. If the linear sequence of SL is subsequently turned into a notional circle, by adding, say, $E < A$ to the sequence $A < B$, $B < C$, $C < D$, $D < E$, the transitive relations are destroyed, but selections in line with original training are still made for the individual adjacent pairs. This can only be explained by adding a contextual factor so that a discrimination can be made between, say, B in the presence of A, and B in the presence of C, as if there were two different B stimuli. Wynne's paper is a demonstration that reinforcement theory can explain seriation without recourse to more cognitive explanations in terms of mental lines or some kind of visual representation of the series against which individual test pairs can somehow be compared, and he urges clarification of theories of this cognitive kind to make them testable against reinforcement accounts.

Acuna, Sanes, and Donoghue (2002) described an elegant procedure for training serial sequences in humans which encourages the formation of some kind of holistic representation (the neurological concomitants of which were the subject of a subsequent fMRI study (Acuna, Eliassen, Donoghue, & Sanes, 2002) discussed below). As evidence for the formation of some kind of reference structure, Acuna and her colleagues noted the increase in response latencies of correct responses during training, rather than the progressive decrease sometimes found.

One potential way to manipulate experimentally the kind of process or algorithm

human participants may come to employ when derived relations are generated, is by varying the instructions given to participants in different groups. Greene, Spellman, Dusek, Eichenbaum, and Levy (2001) made a study of the effects of instructions on accuracy of SL performance, but without reporting response latencies. Only a minimal A-E series was used, which, as we have seen, permits only one valid test of transitive inference (TI). Some participants were specifically told that after training some kind of inference would be tested. These participants performed significantly more accurately on the BD test than those who were not given this instruction, but the differences were slight because performance was close to ceiling for all participants. Uninstructed participants were given a questionnaire to assess the extent to which they had become aware of a hierarchy in the stimuli, but at whatever stage this was administered, scores on this questionnaire were not correlated with the emergence of accurate responding on BD tests of TI. Later Martin and Alsop (2004) used essentially the same posttest questionnaire with an A-G series, though only one-nodal TI (i.e., BD, CE, and DF), not the two- and three-nodal, was tested. The unaware participants were significantly poorer in terms of accuracy of TI, not performing above chance on two of the three tests (CE and DF), and worse, in fact, than pigeons on a comparable task (Von Fersen, Wynne, Delius, & Staddon, 1991). Their similarly poor performance on tests of CD and DE training pairs indicated that the criterion of preliminary training had been set too low in this study to support SL learning in the unaware group. Again, latencies were not reported in this study.

Response latencies were measured in a study of SL using phonologically correct nonwords in an A-G design (Anderson & Dickinson, 2003). Participants in an informed group were told that the stimuli formed a series which they could learn, whereas those in an uninformed group were simply exposed to the contingencies. Both groups were trained to the same criterion of accuracy and then given tests of the original training pairs randomly mixed with tests of transitivity with one, two, or three intervening stimuli. There was no significant difference in accuracy between the two groups: percentage correct in tests of TI was only 67.3% (informed) and 62.7% (uninformed)

with one intervening stimulus, but rose to 100% for both groups with two and three intervening stimuli. There was, however, a significant difference in mean response latencies between the two groups, which was discernible both at the end of training (see Figure 1, top, in which the serial position effect can be seen), and in subsequent transitivity testing (as shown in Figure 1, bottom, where both groups show the symbolic distance effect). At all points, the uninformed participants responded more quickly than the informed. One possible explanation for these differences in latencies is that parsimony of explanation maps on to parsimony of operation: a simpler algorithm based on differential reinforcement values of stimuli hypothesized to be operating in the uninstructed group may have involved less neural processing than putative cognitive structures employed by instructed participants. It would be interesting to look for neurological differences between such groups.

Theoretical accounts of SE such as Imam's (2001) also include efforts to account for the set of empirical phenomena in terms of reinforcement theory. Horne and Lowe (1996, 1997) attempted to explain SE in terms of the subject's pre-experimental ontogenetic reinforcement history. Children have learned to name things, and that the names and the things are interchangeable. Subsequently, by naming the stimuli in an SE experiment, equivalence is delivered as a theoretically trivial by-product. This is a view that provoked considerable peer commentary in this journal.

Our own results have shown that SE is achieved more rapidly (faster reaction times and faster emergence) with nameable icons compared with hard-to-name abstract stimuli (Bentall et al., 1993), though here nameability is confounded with meaningfulness. Teaching common names for the abstract stimuli—the same name for all three stimuli in a particular ABC training set—facilitated SE formation, but this was not found when a different individual name was taught for each of the abstract stimuli. It was also possible, by giving paired-associate training between the names of icons, to effect SE relations between the icons themselves that conflicted with those based upon incompatible trained relations between the icons (D. W. Dickins, Bentall, & Smith, 1993; A. B. Smith, Dickins, & Bentall, 1996). But the fact that names may play a part in or

even be sufficient for the formation of SE relations between the stimuli they name does not mean they are invariably necessary. The naming hypothesis is the mirror image of the idea that the very capacity for SE may be a necessary phylogenetic precursor (or pre-adaptation or exaption) of the (distinctively human) capacity to symbolize in the first place (T. E. Dickins & Dickins, 2001, 2002; Place, 1995). As O'Donnell and Saunders (2003) point out, the increasing number of publications demonstrating SE in mentally retarded individuals who have failed to develop language provides falsifying evidence against Horne and Lowe's (1996, 1997) idea that language is a prerequisite for SE.

Taking a more central view of semantic processes (rather than a peripheralized account in terms of naming behavior), the resemblance between SE structures seemingly indicated by nodal effects, and cognitivist entities, specifically the semantic associative network of Collins and Loftus (1975), has been noted by Fields, Reeve, Adams, and Verhave (1991). This may reflect similarities in the way the trained relations are learned in the first place. The greater facility with which meaningful stimuli are learned and integrated into equivalence classes compared with more "abstract" stimuli (Bentall et al., 1993) also suggests this.

Nor can it be assumed that apparently meaningless materials necessarily bypass semantic processing: nonsense words may activate the same semantic systems of the brain as actual words (Price, Wise, & Frackowiak, 1996). A neat experiment (Galizio, Stewart, & Pilgrim, 2001) which bridges a gap between operant studies of SE and the cognitive phenomenon (Bousfield, 1953) of clustering of semantically related items in free recall⁹ in fact used nonsense syllables as stimuli. Trained relations were established between pairs of syllables using a matching-to-sample (MTS) paradigm. Participants were then required, both before and after equivalence testing, freely to recall as many as they could of the individual nonsense syllables, and the order in which these items were retrieved was noted.

⁹If stimuli belonging to several different semantic categories are presented in random order and the participant is then asked to recall as many of these stimuli as possible, there will be a tendency to recall items in sequences belonging to the same semantic category.

Fig. 1. Top: Serial position effect in serial learning. Latencies of trained responses at the end of training in serial learning as a function of the position of the stimuli on the implicit series in an instructed versus an uninstructed group of participants. Bottom: Symbolic distance effect in serial learning. Latencies of transitive responses after serial learning as a function of number of intervening stimuli in an instructed versus an uninstructed group of participants. (Modified from Anderson & Dickinson, 2003).

The results revealed evidence of significant within-SE class clustering that was particularly pronounced after the equivalence tests.

Another cognitivist procedure which explores semantic and associative linkages, and which has been extensively used in neuropsychology (e.g., Henson, 2003; Rissman, Eliassen, & Blumstein, 2003), is that of priming in

lexical decision making. Two stimuli are presented in succession within critical constraints of duration and interstimulus interval. The first stimulus is known as the prime and the second as the target. Some kind of binary decision is required of the participant, such as the lexical decision as to whether the target is or is not a real word. The response is usually

correct, but it is the speed with which it is made that is the key dependent variable. The response to a real word target will be faster if the prime is a real word semantically related to the target, than if both stimuli are real words but have no salient semantic relation. This procedure has been adapted to the study of SE while measuring participants' EEG activity. Here a characteristic negative component of the event related potential (ERP), the N400 wave, has been shown to systematically follow the presentation of two items that are semantically incongruous. This seems also to be the case if the two items are stimuli belonging to two different SE classes, compared with those belonging to the same class. This procedure has been turned to good effect in the study of SE (Staunton, Barnes-Holmes, Whelan, & Barnes-Holmes, 2003; Warren & McIlvane, 1998).

Relational frames and schemas. An important further point to make is that the participant, in the course of constructing the responses that fit the task's demands, may construe some particular kind of relation between sample and comparison. Relational frame (RF) theorists (Hayes, Barnes-Holmes, & Roche, 2002) designate frames such as equivalence and serial order as high-level operants which are somehow brought to bear in such experiments. Cognitive theorists have applied the somewhat different notion of relational schemas (Halford, Bain, Maybery, & Andrews, 1998).

One wonders about the provenance of such forms of responding. To a sophisticated human participant notions of serial order, such as greater than/less than relations ($>/<$) must be readily available to apply to SL, though the notion of equivalence ($\equiv$) may be less automatically available (unless, in either case, some preliminary instruction or cue is given, such as "one stimulus comes after the other" in SL, or, in SE, that this symbol "stands for" or "is a translation of" or "goes with" that stimulus). Whatever the case with SL, it is not logically compelling, in SE, to select C after A, based upon A-B and B-C training, since the stimuli concerned are not the same as each other¹⁰. Certainly if some kind of rule or relation is construed in order to frame successful responding on the task in

hand, the course of subsequent test performance is likely to be pre-empted.

Experimental instructions. With human participants in the social situation of a psychology experiment, much may be influenced by the expectancy or set an individual participant develops. Traditionally in the operant laboratory some effort is made to keep verbal instructions to a minimum and to bring the behavior under the control of the stimulus variables and reinforcement contingencies the experimenter wishes to apply. The use of examples, however, in which a participant is first shown naturalistic stimuli that are related in some preexperimental way, to illustrate that there will be similar, but arbitrary relations between stimuli in the actual experiment, may entail much implicit instruction. An alternative procedure is to give as far as possible entirely explicit verbal instructions, but how the participant interprets these, and how they influence participants' behavior, needs to be determined.

An effort to expose such effects was made by D. W. Dickins et al. (2002). Participants in one group were shown, prior to MTS training, an instructional PowerPoint presentation explaining in principle what subsequent tests of emergent relations would entail, while the other group had a comparable presentation without such details. In subsequent testing, the number of correct responses on the transitivity tests achieved in a limited time window, in proportion to their performance on tests of trained relations, was significantly higher for participants in the first group.¹¹

Participants' reports. A further question here relates to the dimension which postexperimental debriefing¹² attempts to explore, namely that of the so-called implicit/explicit distinction in cognitive function (Mayes, Gooding, & van Eijk, 1997; Reber, Gitelman, Parrish, & Mesulam, 2003). Neuroimaging

¹¹In laboratory studies with larger groups, similar instructions resulted in significantly faster initial latencies in transitivity in instructed participants, whether instructions were given before initial learning of the trained relations or between training and testing (D. W. Dickins, unpublished data).

¹²Systematic debriefing is difficult to do, but should perhaps always be attempted. A correlation between the results of imaging studies, such as pathways and patterns of structures recruited by tasks of different types, and participants' reports, would greatly strengthen the relevance of such evidence.

¹⁰Just as they are not in the case of a symbol and what it symbolizes (as distinct from an icon).

offers an objective way to provide correlative evidence for such a distinction. Sometimes participants, on debriefing, report that the manner in which they selected the correct comparison on their first tests of transitivity was automatic. Others may report deliberate mnemonic strategies they used, such as visualizing the two stimuli to be linked in a trained relation (e.g., A1 and B1) as located together in some imagined space. Subsequent acquisition of linked relations, such as B1-C1, C1-D1, D1-E1, may involve successively incorporating further stimuli into the same space¹³. Such a visual construct seems to map on to the operational notion of an equivalence class with great facility.

NEUROIMAGING STUDIES OF STIMULUS EQUIVALENCE AND SERIAL LEARNING

Before describing methods and findings of some block-designed fMRI studies, we should take stock of the aims of neuroimaging these behaviors in the first place.

Component processes. How can behaviors involved in SE and SL be subdivided plausibly into component processes of the kind likely to correspond to the actions of the brain? One way is to attempt a behavioral dissection by adopting the principles of chronometrics (Posner & Mitchell, 1967), which depends upon analysis of response latencies, an example of which is given below. Another, which is a methodological necessity for fMRI studies, is to devise control tasks likely to have all but a very few key elements in common with the main experimental task. By subtracting the cacophony of total activations on one task from those on another, regional activity crucial to these key elements may be revealed.

Chronometrics. Response latencies are subject to considerable variance, and can include very long latencies—of multiple seconds duration in the case of SE choices—especially early in the emergence of responses involving transitivity, and particularly with multiple nodes. Very long latencies might also be the consequence of lapses of attention, which might justify treating them as outliers. One

way around this is to express data as speeds (Spencer & Chase, 1996), since taking the reciprocal of the latencies will reduce this variance. Another is to impose a limited hold, so that if the response is not made within some arbitrary limit, the trial is aborted and the next trial ensues. The use of a limited hold raises the issue of the extent to which a participant opts for a particular tradeoff between speed and accuracy on the given task. A limited hold is sometimes imposed for theoretical reasons, (e.g., Imam, 2001), and sometimes for pragmatic reasons, as in the delimitation of the duration of successive trials in an event-related fMRI procedure.

Bearing in mind these methodological caveats, an important assumption that response latency is a function of the number and duration of neural operations required for the task in hand can assist in the analysis of behavior (Posner & Mitchell, 1967). A simple example of this chronometric approach offers an inroad on the analysis of the first steps in the MTS paradigm in SE (D.W. Dickins, 2003). This study employed a nonzero delay procedure; that is, it manipulated the duration of a time interval between the offset of the sample stimulus and the onset of the comparisons. As a preliminary hypothesis, it was envisaged that once trained relations had been established there might be five steps in a participant's performance. These were:

1. As soon as the sample appears, *registration* (perception and recognition) of the sample stimulus.
2. This would immediately evoke *retrieval* of the (by now, well established) specific sample-comparison link, leading to the production of an anticipatory representation of the appropriate comparison, which would be held thereafter in some kind of working memory (WM).
3. When, after the delay, the comparisons appear, *scanning* and *registration* of the comparison stimuli.
4. *Recognition* of the comparison stimulus that matches the representation retrieved and held in WM from stage 2.
5. *Organization* of corresponding *motor response*.

These stages are illustrated in the timelines in Figure 2. In Figure 2a the upper timeline illustrates the onset and offset first of the

¹³Two participants reporting such cumulative visualizations were the exceptional individuals who showed no discernible increase of response latency with nodal distance in the study of SE formation using six 3-nodal ABCDE classes by Bentall et al. (1999, Experiment 2).

sample stimulus (open) and then, after a delay, the onset and offset of the comparison stimuli (hatched). These are cleared away the moment the response is made, so that the length of this hatched area corresponds to the measured response latency (reaction time). The numbered boxes below correspond to the five hypothetical stages, with stage 2 (shaded) representing the time it takes the participant to perform retrieval of the representation of the correct comparison, which in Figure 2a exactly coincides, as it happens, with the experimenter-imposed delay. Since this process is here complete by the time the comparisons appear, the participant simply needs to locate the appropriate comparison which matches the representation, rather like identity matching.

In Figure 2b the situation is shown for arbitrary matching in which there is a zero delay between offset of the sample stimulus and onset of the comparison stimuli. In this case, retrieval of the appropriate representation during stage 2 must take place with the comparisons already on the screen. This takes up time before stage 3 can begin, with the result that, in effect, this time is added to the measured reaction time¹⁴.

Figures 2c and 2d illustrate what might happen in the comparison case of identity matching. Consider a participant, naïve to arbitrary matching, but trained instead on identity matching. Here, stage 2 of the five hypothetical stages does not apply: this is represented in 2c by a black area in the row of numbered boxes, since nothing can occur during the delay¹⁵. After the delay in 2c, or if there is no delay, as in 2d, stage 3 can begin immediately. Therefore, the reaction time in identity matching is predicted to be the same with or without a delay, and similar to that in Figure 2a, for arbitrary MTS in well-trained participants given a sufficiently long delay for stage 2 to be completed.

¹⁴ Stage 2 is hypothesised to begin as soon as stage 1 is complete, which could be before the termination of the sample stimulus. To minimize the extent to which this could occur, in the experiment described sample presentation time was restricted to 0.2 s, which empirically proved sufficiently long to support accurate responding on all the tests.

¹⁵ We ignore here any deleterious effects of a delay which is the usual target of experiments in which the delay in delayed MTS is varied. The prediction of the present experiment is in the opposite direction to any such effects.

As an alternative to the anticipatory strategy implied by the immediate initiation of a stage 2 retrieval process once the sample has been registered in stage 1, participants could adopt a retroactive strategy of waiting to see the comparisons before attempting to retrieve the appropriate trained relation¹⁶. Here, no benefit would accrue in terms of a faster response time after a finite delay compared with a zero delay. This is illustrated in Figure 2e, in which the black box again corresponds to the absence of any stage 2. During the delay the participant simply holds the sample in WM and awaits the comparisons. The ensuing scan of the comparisons combined with a symmetric search for each to retrieve a representation of the sample it had been paired with in training, and to check whether this matches the representation of the sample on this trial, might take longer than the hypothesized stages 3, 4, and 5 in the anticipatory strategy.

It was found (D. W. Dickins, 2003) that in tests of performance of well-trained arbitrary relations, response latency was a negative function of increasing delay, whereas in identity MTS (in the same participants, prior to training on arbitrary MTS), the same range of delays had no effect upon response latency (see Figure 3). From the group data, the latencies for arbitrary MTS at zero delay were on average 0.3–0.4 s longer than the mean response latency of 0.9 s for identity matching, suggesting that stage 2 took on average between 300 and 400 ms. If, however, response latency is plotted against increasing delay values, the latency progressively declines until a delay of at least 1 s is imposed. This mismatch may be due to the fact that as more and more tests were conducted, participants continued to benefit from cumulative practice on stages 3, 4, and 5 and became faster on the matching process itself, and/or that the matching process in identity differs in some respects from that in arbitrary matching, the former being based more on comparing a number of perceptual features and the latter upon semantic coding (see discussion on use of identity matching as a baseline).

The same effect of accelerating response speed by increasing the delay, up to different

¹⁶ Similar “prospective” and “retrospective coding” strategies have been described in pigeons (Grant & Kelly, 2001).

(a) Arbitrary matching: anticipatory strategy: finite delay - trained relation retrieved before onset of comparisons

(b) Arbitrary matching: anticipatory strategy: zero delay – retrieval of trained relation while comparisons on screen

(c) Identity matching: finite delay

(d) Identity matching: zero delay

(e) Arbitrary matching: retroactive strategy: finite delay - trained relation retrieved only when comparisons appear

Fig. 2. Timelines for arbitrary MTS compared with identity matching with zero and nonzero delays (D. W. Dickinson, 2003) (not drawn to any particular scale). Upper lines: onset and offset of sample stimulus (open) and comparison stimuli (hatched), with or without an interstimulus interval (nonzero or zero delay). Response latency is measured from the onset of the comparison stimuli until the response. Lower boxes: postulated succession of hypothetical component processes. Key: 1 = registration, 2 = retrieval of trained relation, 3 = scan of comparisons, 4 = recognition of correct comparison, 5 = organisation of response; black unnumbered box indicates no postulated process occurring. See text for explanation.

Fig. 3. Effects of delay on identity matching and arbitrary (i.e., symbolic) MTS tasks. Shown are the mean of the median response latencies after zero and nonzero delays between sample offset and onset of the six comparison stimuli.

optimal delay values, was found also for subsequent tests of symmetry, transitivity, and equivalence in the same participants (D. W. Dickins, 2003; D. W. Dickins & Kramo, 2003), providing evidence for stage-2 type anticipatory processes in derived responses as well as in trained relations.

The subtraction paradigm. The cacophony of activity in the brain as a whole continues unabated in waking and in sleep, and local changes in the rivulets of this cascade can only make sense by appropriate techniques of statistical analysis and of modeling. The most commonly employed starting point is to employ the subtraction paradigm, which rests on the assumption that there is functional specialization in the brain: that is, that different regions are engaged in different computations. The subtraction paradigm is implemented using various statistical techniques that examine images voxel-by-voxel (e.g., Friston et al., 1995). By such means, systematically repeated events can be discerned when one behavioral task is compared with another closely similar one, devised to have all but a very few key elements in common. Contriving these contrasts is the skill of the neuropsychologist, and it has to be, to a large degree, an empirical matter. Some subtractions used in the study of derived relations and related fMRI studies are reviewed below.

The locality of activations isolated by the subtraction paradigm, typically in terms of significant clusters of differentially active voxels, is the chief dependent variable fMRI yields, and if a number of sites appear to be

jointly implicated by such contrasts, then we are challenged to construct an outline functional story about what these bits of the brain do. A problem here, however, is that the role of many other regions, doubtless making vital contributions to the common elements of the experimental and control tasks, get lost in any one such comparison. To come to grips with such functional connectivity, and detect possible functional networks, is the task of the covariance paradigm, which depends upon correlation and regression analyses, and more complex multivariate methods of analysis.

Beyond these descriptive researches lies the domain of neural modeling that actively hypothesizes how such functional networks operate in the brain. The interested reader is referred to an introductory review of this topic by Horwitz, Tagamets, and McIntosh (1999).

AN fMRI STUDY OF SE

An example of the logic of subtraction was used in a block design by D. W. Dickins et al. (2001). This was a study of the emergence, in the scanner, of six potential ABC equivalence classes using readily nameable iconic stimuli taken from the Macintosh Mobile font. Training of the six AB and six BC relations was given in a conventional laboratory on a previous day, using a zero delay, arbitrary MTS procedure. For example, in AB training, one of the six A stimuli would be presented for 2 s as the sample stimulus. Offset of the sample would be followed with a zero delay by presentation of a row of the six B stimuli as comparisons. The response was a key or button press with the appropriate finger from the three middle fingers of each hand corresponding to the position of the chosen stimulus in the row. Any response terminated the comparisons screen and activated an appropriate feedback screen which indicated whether or not the response was correct. During training the participant also was required to make another response which resulted in the presentation of a new sample constituting the beginning of the next trial.

When participants had reached criterion on the trained relations, they were given a practice test run to prepare them for the tests to be conducted in the scanner. No feedback or knowledge of results was given on test trials.

The run lasted 5 min and was divided into twenty 15-s periods, with 10 periods in the *off* condition alternating with 10 in the *on* condition. For the practice run and all the runs in the scanner, the *off* test trials consisted of a simplified form of identity matching. The identity matching had an asterisk stimulus as sample, and an identical asterisk as the single comparison stimulus, the other five comparison boxes being empty. The *on* task in the practice test run consisted of test trials of intermixed A–B and B–C trained relations.

Within any one period during a test run, trials were presented as in training, except that there was no feedback screen, and onset of the next sample now occurred automatically, 1 s after each choice response. Trials continued like this, and participants worked at their own pace, until the 15-s period came to an end. At this point overriding control was assigned to the clock, so that switching between the *off* and *on* periods took place without regard to the stage in the cycle of trials or the participant's behavior. There was no indication that this disrupted performance: judging from the subsequent speed and accuracy of behavioral performance over the entire scanning period which entailed four such 5-min runs (see below), all subjects were significantly faster and more accurate on the *off* task, and performance was sustained, or even improved slightly, over the entire period. In addition to its fMRI functions, the *off* task therefore provided a control for any decrement in motivation, accumulation of fatigue, or discomfort, from being in the scanner. Participants who had in any way indicated distress or discomfort in the scan were allowed to withdraw from the experiment immediately, treated with sympathy and respect, and given the same expenses payment (£15) as their fellow participants.

In the scan, subjects were given four, 5-min blocks of test trials. In each block, ten 15-s *on* episodes alternated with ten 15-s *off* episodes. For the successive blocks, the *on* task consisted of test trials of trained relations, symmetry, transitivity, and equivalence respectively.

For fMRI purposes the *off* task was designed to involve peripheral aspects of the execution of MTS, which could be discarded by subtraction from the *on* task, thus leaving all of the more central components relevant to the

computation of SE open to exposure in this first investigation. The resulting activations provide a starting point for future research. But what exactly were the components that were discarded in this design?

It is very likely that each participant soon learned to associate presentation of the distinctive asterisk stimulus as a cue for the ensuing test, which simply entailed locating the occupied comparison box and executing the appropriate response. In comparison with the experimental tasks, the load in the control task on working memory (WM) would be slight, and little perceptual comparison would be required. Motor output would be essentially the same in the control as in the experimental tasks.

Brain activation results. Based upon 100 acquisitions of the echo-planar images for each 5 min block, regions of the brain showing a hemodynamic response significantly correlated with the on/off profile of the task were identified. As these activation data were spatially normalized, both single-subject and group statistical analyses could be performed.

Principal regions activated in all 4 test blocks were bilateral dorsolateral prefrontal cortex (DLPFC) adjacent to the middle and inferior frontal gyri (viz., Brodmann's areas (BA) 6—except in failed transitivity participants, see below—and BA 9, 44, 45, 46); bilateral posterior parietal cortex (BA 7, 39, 40); medial frontal gyrus (BA 32); anterior cingulate cortex; and bilateral primary visual cortex (i.e., calcarine sulcus).

Unlike the patterns in the same participants in the same scan on a test of verbal fluency, no activation of Broca's area (more ventrally in BA 44/45) was found in any of the MTS tasks. Such activation has been associated at least as strongly with silent as with overt speech (Huang, Carr, & Cao, 2001). Although a total lack of Broca activity during SE cannot be incontrovertibly inferred simply from a qualitative comparison between statistically significant subtractions performed within two different tasks, even when performed by the same participants in the same scanning run (James Pekar, personal communication), these results give no support to the simple behaviorist view that mediation of stimulus equivalence relationships occurs when the subject either overtly or covertly says the names of the stimuli (Horne & Lowe, 1996, 1997). Future fMRI

studies directed at this specific issue may confer the decorum of Popperian falsifiability upon this otherwise elusive naming hypothesis.

Because of a dichotomy of participants' scores on transitivity (7 participants 66–96% correct, 5 participants 8–33% correct), it was possible to make comparisons between participants' activations on this task. The successful subjects demonstrated a higher degree of left lateralization in DLPFC compared to those who failed. Furthermore, activation in BA 10 was bilateral in the unsuccessful group and only present on the left in the successful group. Caudate activation was present only in the successful subjects and again, this was left lateralized. In addition, identical patterns were found in the data for the trained relations condition when subjects were allocated to the same subgroups and their data reanalyzed.

Individual subjects' frontal lobe lateral indices¹⁷ were compared with their performance on trained relations, symmetry, transitivity, and equivalence. There were significant correlations (Pearson's product moment) between the degree of left lateralization of activity in the frontal cortex and the percentage accuracy of responses for trained relations ($r = -.724$, $p < .05$), transitivity ($r = -.790$, $p < .01$) and equivalence ($r = -.793$, $p < .01$), but not for symmetry. Lateralization of frontal activity also correlated significantly with response latencies in the tests of transitivity ($r = -.930$, $p < .01$) and equivalence ($r = -.655$, $p < .05$), with the more left-lateralized subjects responding more rapidly. No corresponding relationships were identified for parietal lobe activation.

It is difficult to be sure what these lateralization results mean. There is a profusion of arguments in the literature for and against various specific roles for left and right prefrontal cortical areas (see e.g., Lee et al., 2002; Rushworth & Owen, 1998; Shallice, 2003). On the one hand, it is tempting to speculate that some left-lateralized verbal semantic process crucial to SE is being more effectively deployed in the DLPFC of more successful participants, either during some process of transitivity, or at a more fundamental level in

the manner in which the original trained relations were learned. On the other hand, it may be that the greater activity in the right DLPFC in the less successful participants indicated the greater difficulty for them of matching to sample, both trained and emergent, requiring the recruitment of a larger number of voxels from this region and thus more bilaterally. Also, transitivity would be expected to be harder if the baseline trained relations had not been sufficiently overlearned.

In a second fMRI study of SE (D. W. Dickins et al., 2002), a different control task was introduced. The task was full identity MTS, with AA, BB, and CC trials, that employed any A, B, or C experimental iconic stimulus as the sample (chosen in randomized counterbalanced fashion for each trial), together with all six stimuli from the same array (either A, or B, or C) as comparisons, which thus included a reiteration of the sample stimulus, which was the one to choose. This task can be expected to differ from the asterisk control in a number of respects. Unlike the asterisk, the sample stimulus presented gave no intrinsic indication as to the upcoming trial, although the experimental design entailed the successive presentation of up to five such identity trials in each 15-s period (like the periods described above), making them reliably predictable until a change ensued at the onset of the next period. Given the zero-delay design employed, which ensured that sample and comparison stimuli were never on the screen together at the same time, for the purposes of identity matching the sample stimulus would need to be held briefly in some kind of store, such as an afterimage or an iconic store. On experimental trials, in which the appropriate comparison stimulus was always a different stimulus from the sample stimulus, it can be hypothesized that a further kind of representation of the sample (a neural holding process or WM) would be needed, pending the selection of the appropriate comparison stimulus.

In full identity matching, unlike the asterisk control task, some kind of perceptual search among the field of six comparisons would be necessary to locate the comparison that matched the sample most closely. Although the search may have been in terms of perceptual features, and predictably simpler

¹⁷ Lateral index (LI) may be expressed as the ratio $R-L / R+L$, where R and L = right and left activation, respectively. Thus, for completely unilateral right activity, $LI = +1$, and for completely unilateral left activity, $LI = -1$.

than a conceptual match on the experimental trials, it would be expected to involve more processing than simply locating the occupied box as in the asterisk task. Ideally, this higher-level control task might be expected simply to add some extra components to the list required by the lower-level asterisk task, involving specifically the memorial and perceptual requirements of registering both the sample and the comparison stimuli. Activations associated with these could simply be subtracted from the activations of the experimental tasks to clarify the more central operations of retrieval of paired associations, with or without the necessity to access representations of nodal stimuli not presented during the trial, as in tests involving transitivity. The last, nodal function might be exposed by using tests of trained relations as the control, with transitivity or equivalence as the experimental task. These assumptions of simple additivity lay behind the actual experimental design of D. W. Dickins et al. (2002). The results, however, implied a Venn diagram more like that of Figure 4b than the additive model of Figure 4a.

Comparisons were made between three different tasks in each of three runs. In the first run, activations for the asterisk control task used in D. W. Dickins et al. (2001) were compared with those for the full identity control and with those for tests of trained relations. Fifteen-second periods of each of the three tasks followed one another in the same order 10 times, making a total run length of 7.5 min.

The major clusters of activation emerging from the various subtractions afforded by this triple design are itemized in Table 1, which gives the numbers of voxels in each cluster; the stereotaxic coordinates (Talairach & Tournoux, 1988); and the name of the brain region together with the Brodmann area in brackets. Highlights of the comparisons are shown schematically in Figure 5.

Triple comparisons: the three baselines. Behaviorally, reaction times (time between onset of the comparison stimuli and the response) support a hierarchical comparison between the three tasks, like that in Figure 4a. The means of the median latencies on Run 1, for 9 participants, were 1.87 s for trained relations, 1.27 s for full identity (ID) matching, and 0.84 s for the asterisk task: only 1 participant

showed a reaction time for ID as long as that for arbitrary MTS, with all the others showing the systematic three-step relation. These differences between identity matching and trained relations were consistent for all 9 participants across all three runs of the experiment.

What is clear from the details of the triple comparisons in Table 1 is that the full identity task, using the full set of iconic stimuli and a full field of six comparisons, entailed considerably more parietal and frontal activity than the simplified asterisk control. The greater activity in the left occipital gyrus also indicates that greater perceptual analysis was required. For this task it seems likely that a visual image is retained and matched against the comparisons directly.

Arbitrary MTS can be expected to involve more operations than identity MTS because the sample must be held in some kind of memory while it is checked against a record of trained relations. This may involve transmutation of the representation of the stimuli into some kind of verbal or semantic code (see discussion below of the work of Sevostianov et al., 2002). Whether this retrieval and comparison process starts during the time when the sample is on the screen, or whether the presence of comparisons, both the correct comparison and the foils, triggers the necessary search, cannot be determined from this design.

However, when the trained relations task is compared with full identity matching by subtracting the latter from the former, increased activity is indicated only in two clusters in the left parietal, together with a cluster in the left superior temporal gyrus (BA 39). Interestingly, this result implies no more work for the frontal areas (DLPFC). Indeed, if the subtraction is done the other way round, several areas involved specifically in full identity matching become apparent. These in fact include frontal areas, the medial and inferior frontal gyri (BA 9 & 46) on the right, but with a small involvement of 20 voxels of middle frontal gyrus (BA 10) on the left. Not surprisingly the same left occipital activity seen in the comparison between full identity matching and the asterisk task is again found, but also some parietal activity on the right (BA 40) and even on the left in BA 7, which must correspond to a different subdivision of this region than that implicated in the inverse

Fig. 4. Schematic representation of shared and specific features of the three types of matching-to-sample tasks employed in D. W. Dickins et al. (2001). The results of this fMRI investigation conformed more closely with (b). (a) An additive hierarchy in which the upper tasks encompass all of the features of those below them, plus specific features of their own. (b) A Venn diagram in which each task shares some features with one and some with both of the other tasks, but also has its own peculiar features.

Table 1

Principal clusters of activation emerging from paired comparisons between the three tasks in the first run of Dickins et al. (2002). The second named task was subtracted from the first in each comparison. (ID = Full Identity matching; ASTERISK = the simplified identity matching control task.) (See text for further explanation.)

Cluster (number of voxels)	Talairach coordinates			Brain region (Brodmann area)
RUN 1			9 Participants	
ID vs ASTERISK				
76	-20	-61	44	Left parietal lobe (BA7)
144	6	17	48	Right superior frontal gyrus (BA8)
65	32	-67	50	Right parietal lobe (BA7)
42	-35	-73	-13	Left occipital gyrus (BA19)
54	-47	21	18	Left inferior frontal gyrus (BA45)
Trained relations vs ASTERISK				
216	-20	-73	47	Left parietal lobe (BA7)
27	-50	24	29	Left middle frontal gyrus (BA9)
13	-47	9	24	Left inferior frontal gyrus (BA9)
Trained relations vs ID				
82	-32	-56	30	Left superior temporal gyrus (BA39)
28	-20	-71	36	Left parietal lobe (BA7)
14	-53	-47	46	Left parietal lobe (BA40)
ID vs trained relations				
54	47	32	9	Right inferior frontal gyrus (BA46)
20	-41	52	-5	Left middle frontal gyrus (BA10)
60	-17	-55	55	Left parietal lobe (BA7)
33	-35	-73	-13	Left occipital gyrus (BA19)
166	8	39	31	Right medial frontal gyrus (BA9)
91	38	-49	49	Right inferior parietal lobe (BA40)

subtraction of trained relation minus identity. These activation differences, therefore, suggest an interrelation between these three tasks more like the Venn diagram in Figure 4b.

Our studies adopted a block design, which meant that activation associated with all the events within and between trials in any one period were being lumped together, despite variations in the number and speed of trials which is determined partly by the participant's behavior and thus confounded with type of trial.

Though also within a block design, an elegant and relevant comparison of tasks was carried out by Sevostianov et al. (2002), who scanned participants performing on three different conditions on a "one-back" task. In this task, stimuli were presented sequentially, and for each new stimulus the participant's task was simply to report whether the stimulus was the same as the previous stimulus, or different. Either pictures, or words, or a combination of the two, were used. This was achieved by having schematic drawings of living and nonliving objects as the pictures, and the printed names of the same objects as the words. The three conditions, with which participants had become familiar in prelimi-

nary training, were all pictures (P), all words (W), or the mixed condition (M), in which words and pictures alternated. Condition P has strong resemblances with full identity matching in our own study, whereas condition M more nearly resembles tests of trained or emergent relations if it is the case that the visual stimuli, particularly the nameable icons used in these experiments, need to be recoded semantically in verbal terms in order to access the trained relations to derive the correct response. Although condition W might be thought of as being more like simple identity matching (and participants reported they adopted a similar strategy for both P and W, unlike that for M), the imaging results suggested that the verbal stimuli in fact triggered semantic networks even if these were not contributing to the selection of the appropriate response.

Contrasts formed included those when P was compared to W ($P > W$), W compared to P ($W > P$), and the mixed condition compared to P and W ($M > P + W$), and the results of these comparisons are summarized below.

$P > W$. Two large areas of activation were found in right and left occipitotemporal cortex (BA 19, 37), but three times greater in

Fig. 5. Glass brains showing significant differences when activations obtained on various control and experimental tasks were compared (Dickins et al., 2002. On the right of each pair is a transverse section, and on the left is a sagittal view of a transparent brain, with the anterior pole to the right. Dark areas indicate significant clusters of voxels, and the salient ones are labelled with their Brodmann areas. (a) full identity matching task minus asterisk control task; (b) trained relations task minus full identity task; (c) full identity task minus trained relations task; and (d) trained relations task minus asterisk control task.

the right than in the left. This right bias accords with other literature (cited by Sevostianov et al., 2002) on the representation of visual objects in the human ventral visual pathway, especially if the objects were not explicitly named (Beason-Held et al., 1998; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Haxby, Ungerleider, Horwitz, Rapoport, & Grady, 1995; Ishai, Ungerleider, & Haxby, 2000). Sevostianov et al.'s participants reported in exit questionnaires that the one-back P condition was the easiest to do because they could directly compare the sequentially presented pictures without explicitly identifying each object, suggesting performance primarily at the perceptual level.

$W > P$. In the W task participants indicated that they had just compared letter strings, and did not convert the words into their meaning. The three left temporal activation clusters from this subtraction appear at first to cast some doubt on this latter claim. The activations were in the fundus of the superior temporal sulcus (STS), the middle temporal gyrus (MTG, shown by others to be activated for word and word-like stimuli; Price et al., 1996; Rumsey et al., 1997), and the Sylvian fissure (BA 42/40). There is, however, evidence that verbal and verbal-like stimuli acquire access to the lexicon automatically (Price et al., 1996), a point to bear in mind when using nonsense syllables. Also, activation was found neither in Broca's area (BA 44/45), where signs of any verbal rehearsal might have been expected, nor in Wernicke's area. Such results suggest that extensive semantic processing may have been unnecessary.

$M > (W + P)$. In the mixed condition the requirement to relate pictorial to lexical representations necessitated a decision at the semantic level, and participants considered this task the most difficult. In this contrast the two largest clusters of activation were obtained along the left precentral sulcus, extending from BA 4/6 rostroventrally to BA 8/9, and in right MTG / STS, plus some smaller clusters in the right precentral sulcus and in medial frontal cortex (BA 8). The extensive activation of left DLPFC (dorsolateral prefrontal cortex, BA 8/9) conforms with its role in the literature as a central executive (e.g., Fletcher & Henson, 2001; Postle & D'Esposito, 2000; E. E. Smith & Jonides, 1999). In the mixed condition, component subtasks must

include coordination of picture and word recognition, retrieval of semantic information from long-term memory, and monitoring of conflicts when word and picture do not match.

Such mismatches were scheduled for half the trials, which may explain the medial frontal activation found inside the interhemispheric fissure, along with the bilateral clusters found in the dorsal precentral sulcus. There is a Stroop-like¹⁸ inhibition at the semantic level between nonmatching pictures and words (Glaser & Glaser, 1989), and a similar pattern of activations in dorsal BA 6/8 was reported by V. Menon, Adelman, White, Glover, and Reiss (2001) for a Go/NoGo response inhibition task, and for incorrect equations in arithmetic (V. Menon, Mackenzie, Rivera, & Reiss, 2002).

The second largest cluster of activation in the $M > (W + P)$ comparison was obtained in right MTG / STS / STG, an area frequently activated during silent word generation (Indefrey & Levelt, 2000). Just as words have privileged access to the lexicon, so pictures have privileged access to the semantic network. Since words are processed more automatically than pictures (Greenham, Stelmack, & Campbell, 2000), the functional linkage between pictures and words required in the Mixed task would necessitate greater activity in the semantic module than in the lexicon, and this right bias would be predicted.

fMRI STUDIES OF SL

As we have seen, SL resembles SE in that it involves the derivation of novel relations from a number of trained relations, in this case simultaneous discriminations in which a particular stimulus (say B) is the correct one to select (+) when it is paired with one stimulus (A) but not correct (−) when it is paired with another stimulus (C). If these pairs form a linear sequence or hierarchy, set out schematically as follows,

A – B +, B – C +, C – D +, D – C + ... etc.,

¹⁸So-called because it resembles the delay in reporting the color of the ink in which a word is written when the word is the name of a color incompatible with the ink color (Stroop, 1935).

then novel choices, between stimuli which have not been paired in training, will tend to be made systematically in conformity with their relative positions in this sequence. Unlike transitivity in SE, these transitive inferences (TI) are faster, and sometimes more accurate, the more intervening stimuli there are between the two stimuli in the test pair (the Symbolic Distance Effect, in contradistinction to the Nodal Distance Effect in SE). There is also a Serial Position effect in SL, not applicable to SE: pairs involving stimuli nearer one end or other of the putative series will be processed more quickly than those involving stimuli from the middle.

Metaknowledge of this sequence may not be necessary for such choices, nor is it easy to establish, even in participants in whom such metaknowledge can be demonstrated, that it may in some manner determine such choices. Unlike SE, SL can (arguably) be achieved by nonhuman species, including rats and pigeons as well as primates; and general learning theory—not a specific theory concocted simply in order to explain the results of SL studies in animals—may provide a satisfactory account of these phenomena (Wynne, 1995). There is some evidence from differential response latencies (Anderson & Dickins, 2003) that humans may achieve TI in more than one way, depending upon the instructions they have been given during training. Imaging studies would seem to offer a way to distinguish between possible different levels of explanation of SL, perhaps akin to the different patterns of brain activation associated with approximate and specific arithmetic calculation (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999).

Acuna, Eliassen, Donoghue, and Sanes (2002) performed an fMRI study of SL following up an elegant laboratory study of seriation (Acuna, Sanes, & Donoghue, 2002) in which participants were trained on pairs of discriminations derived from a series of 11 stimuli differing in an unrelated way both in shape and color, and presented in random order. Acuna, Eliassen, et al. (2002) scanned behavior on what is, for seriation, paradoxically the hardest test of transitive relations, that of the non-adjacent-1 relation, in which two nonendpoint stimuli from a series have just one other stimulus in between, such as B vs. D, or E vs. G in their A→K series. On this test of

TI, participants were found to be slightly but significantly both slower and less accurate than on tests with nonadjacent stimuli with more intervening stimuli, in accordance with the symbolic distance effect.

In the scan, TI was compared with a control task of height discrimination. Here participants were given adjacent stimuli from the A→K series, but were required to choose the taller of the two stimuli, irrespective of their serial position. This they could do completely accurately, and even faster than the endpoint adjacent discriminations A vs. B and J vs. K in the SL task. Thus, this control task involved visual processing of the same stimuli, and the application of a rule for discriminating between them, but without the necessity of referring back to a file of remembered relations derived from past trials, decision processes, and button pushing. The height test (HT) was in turn compared with a further task (VIS), designed as a control for purely visual processes, in which participants were required to fixate a central point between pairs of stimuli that were similar to those in the SL task, but were not members of any series, without making a response.

Brain activations from the TI-HT comparison were, in sum, as follows. For the TI-HT subtraction, the right superior frontal gyrus (BA 6) showed a medial and a lateral cluster of activation; there was bilateral activation in the middle frontal gyrus; on the right in BA 8, 9, and 46; on the left in BA 6, 8, 9, and 46; and in the left inferior frontal gyrus, BA 8 and 9. In the parietal region there were large clusters bilaterally in BA 5, 7, and 40.

For the HT-VIS subtraction, frontally there was bilateral activation in the precentral gyrus (BA 4 and 6) and in the medial frontal gyrus BA 6. Parietally, there was some left lateralized activity in BA 7.

Thus, there were broad similarities with our findings for SE (D. W. Dickins et al., 2001; D. W. Dickins et al., 2002) involving prefrontal cortex in conjunction with bilateral parietal activity associated with what Acuna, Eliassen, et al. (2002) see as the application of some rule and subsequent computation of TI.

For Acuna, Sanes, and Donoghue (2002), the implication of the results on their SL paradigm is that participants build some kind of representational structure during the learn-

ing of the component conditional discriminations. This they say is evidenced by the increase in response latencies as learning progresses in all but the endpoint comparisons. This increase occurred at the same time as the variance of these latencies diminished and the proportion of correct responses increased. The participants were told they were learning a series, and a criterion of adequate training (prior to testing) was that they were able to sketch the stimulus series from memory, although Acuna, Sanes, and Donoghue do not say how accurately participants were required to replicate the shapes and colors, nor how well they did. Nor is it clear what the cognitive structure might be, nor how it might be used. It is not entirely clear how the serial position effect, and the effect of proximity of discriminated pairs to either end, or particularly the K end, of the A→K series, which Acuna, Sanes, and Donoghue demonstrated, can fit into such a model. Discrimination of some specific value or quality, developed through some history of differential reinforcement, and attached to a different degree to each individual stimulus or to each stimulus in a given context, would account equally well for these effects (Wynne, 1995).

Whether participants might be able to make correct transitive responses using the same procedure, but without explicit mention of an overarching series embracing the individual conditional discriminations (as appeared to be the case in a similar experiment by Anderson & Dickens, 2003), deserves to be investigated, including whether this is possible without some consequent mental line or other representation; and if so, whether different brain regions might be involved. Such a possibility seems likely to be true for nonhuman subjects that solve SL problems. Similarly, in SE, different degrees of explicit instruction given to participants could be explored systematically, and the capacity of participants to generate equivalence classes (e.g., by sorting the stimuli into groups after training and before testing, or after testing) should be investigated. It might be the case, too, that different participants solve these problems in different ways.

A recent fMRI study of SL by Heckers, Zalesak, Weiss, Ditman, and Titone (2004) links results reportedly similar to those of Acuna, Eliassen, et al. (2002) to an interesting

analysis of the role of the hippocampus. Only a 5-member A-E series of stimuli (visual patterns) was used and so the only valid test for TI was that of novel BD pairings. However, tests anchored¹⁹ at one or the other end of the series, namely, AC, AD, BE, and CE, also were given. A further control involved simultaneous training on four nonoverlapping pairs of patterns, characterized as ab, cd, ef, and gh, and testing of five novel conjunctions of these, namely ad, af, cf, ch, and eh, which were, in fact, anchored at both ends.

In terms of accuracy, Heckers et al. (2004) found no behavioral differences in performance on these tests, but response latencies did show significant differences, with faster responses for non-overlapping pairs than for overlapping pairs, both for previously trained and novel combinations. The overall pattern of brain activation they describe as similar to the findings of Acuna, Eliassen, et al. (2002). Activation was found in a network of brain regions that involved pre-SMA (Brodmann area 6), and bilateral parietal, prefrontal, and inferior temporal cortices. This pattern, however, was found with non-transitive inferences also, and with any test involving overlapping pairs of stimuli. Various subtractions (including a comparison between BD and the other novel but anchored tests of the overlapping series) converged on implicating right anterior hippocampus only on TI tests (i.e., in those in which relational judgements had to be made), whereas the parahippocampal gyrus was involved with performance on non-relational tests.

It is worth reviewing briefly two other recent studies, one a PET study in humans and the other an experimental surgical study in monkeys, which support a coherent story about the role of the hippocampus in deriving new relations. These studies are relevant to our forthcoming study of the emergence of SE which is described briefly in the final section.

Nagode and Pardo (2002) conducted a PET study of both learning and testing in which participants were given both an overlapping and a nonoverlapping sequence of photographs of faces. They found that hippocampal activation did not occur until participants

¹⁹ Anchored: by including an end stimulus, the choice of which in training had been either always reinforced or never reinforced.

reached an intermediate stage of training on the overlapping pairs. This time course of brain activation matched participants' self reports that they were not consciously aware of integrating the faces into a hierarchy until they had actually memorized several separate face pairs. By the end of training, when the individual overlapping pairs had been fully learned and presumably the overlapping information had been integrated, hippocampal activity returned to its baseline level. The results indicate that thereafter, on tests of TI, only a "trivial readout" is required to derive the answer from previously developed knowledge about the hierarchy. Nagode and Pardo relate their findings to the model proposed by Eichenbaum, Otto, and Cohen (1994) in which there is a division of memory labor between the hippocampus and its surrounding cortices whereby the parahippocampus encodes simple associations, whereas the hippocampus integrates them.

A recent study (Buckmaster, Eichenbaum, Amaral, Suzuki, & Rapp, 2004) using three cynomolgous monkeys (*Macaca fascicularis*) that had received bilateral entorhinal cortex ablations before training on a 5-member SL series found that the monkeys performed below chance (despite scoring 73.3% correct) on BD tests of TI, and significantly less well than the four unoperated control monkeys (that scored 100%). The lesioned monkeys nevertheless continued to perform as well as controls on the AE end-anchored tests in the A-E series, and on other tests based on memory for individual stimuli (delayed non-matching to sample and object discrimination learning), but reportedly showed deficits on some other tests "that emphasized the flexible manipulation of memory for the relationships between familiar items" (besides TI, a paired-associate task and a spatial delayed-recognition-span procedure).

FORTHCOMING STUDIES

At the time of writing I am collaborating on two new prospective fMRI studies of SE. Both involve a trial-by-trial analysis by employing event-related designs (Josephs & Henson, 1999). These designs enable events occurring within the confines of specific trials to be averaged without inclusion of intertrial events, incomplete trials, or other irrelevant events that

necessarily occur in a whole block of trials, thus greatly sharpening comparisons. Also, the types of trial can be allocated and analyzed post hoc, as, for example, when distinguishing between correct responses and errors. Furthermore, by jittering the onset of individual trials, that is varying the phase of the sample stimulus onset relative to that of the fMRI acquisition, the time course of the BOLD response associated with a particular kind of trial can be constructed (Josephs, Turner, & Friston, 1997).

One of these studies, in Baltimore, is being conducted in collaboration with Lanny Fields, Mike Schlund and Mike Cataldo, and the other is in Liverpool with Andrew Mayes, Daniela Montaldi, and Neil Roberts in collaboration with Dermot Barnes-Holmes. In both studies the stimulus-pairing yes/no procedure (Fields et al., 1997) or Relational Evaluation Procedure (REP) (Cullinan et al., 2001) is being used as the behavioral paradigm for SE, which should provide useful comparability between the two sets of studies. The studies differ mainly in (a) types of stimuli employed, and (b) the tasks used as controls in subtraction analyses.

The Baltimore study consists of a series of event-related experiments using nonsense syllables as stimuli. Participants will be trained on two ABCDE equivalence classes that contain 1-, 2-, and 3-node emergent relations. Training of baseline relations and testing of emergent relations will be done in the laboratory prior to scanning. Control trained relations that do not share potential nodal links, such as WX and YZ, also will be taught to provide for pseudo-transitivity tests such as WZ and YX, and pseudo-equivalence tests such as ZW and XY. After SE class formation, the neural concomitants of nodal function will be extracted by comparing the activation patterns occasioned by transitivity probes and equivalence probes with those occasioned by pseudo-transitivity and pseudo-equivalence probes given in the scanner. The use of multinodal classes will enable us to identify differences between activations as a function of the number of intervening nodes. In another experiment, scanning will occur during initial presentation of stimulus pairs prior to any training. Training and SE testing then will be given, and the same regimen of stimulus pair presentation as in the initial scan will be performed again in a second scan of the now

experienced participants. Differences between the patterns of activation for the first and second scans for each individual will be assessed to determine any systematic changes for each stimulus pair once it has become a relation in an equivalence class.

The Liverpool study takes a different approach by using meaningful stimuli falling into different categories, for example photographs of strangers' faces, written words, pictures of familiar objects. This use of meaningful stimuli has numerous consequences, but the primary purpose is to exploit the discriminably different loci in the brain known from the literature to be involved in processing stimuli of different categories (e.g., for faces, Henson et al., 2003; Kanwisher, McDermott, & Chun, 1997; for faces versus objects, Haxby et al., 2001).

A preliminary fMRI investigation, employing a block design, and not involving SE, will be conducted to identify the patterns of activation evoked by these different stimuli in the context of the Relational Evaluation procedure (REP), and, in particular, by the word stimuli (animal names) that will be deployed as nodes in the complementary event-related study to be described below. In the scanner the boxcar design will consist of consecutive 30-s periods of three different tasks presented in counterbalanced random order six times, making a total run length of 9 min. One task will consist of six REP test trials of trained object-face relations. The second task also will test trained relations, in this case between a second set of objects paired with words. On each of the six trials only the object stimulus will be shown, and participants will be instructed to recall (silently) the word with which this object was paired in training, and to make a choice response on the *same* button if they thought they could remember the correct word, and on the *different* button if they could not. The third task will be a baseline task consisting of six simultaneous discrimination trials between two adjacent visual patterns which may be identical or may have subtle differences between them, again requiring an appropriate *same* or *different* response.

The event-related study employs SE training with an AB, BC design, involving the potential formation of a number of ABC equivalence classes. 'A' stimuli are faces, 'B' are the written

names of familiar animals, and 'C' are pictures of common man-made objects. Alongside the pre-scan training of interconnected AB and BC relations a further series of direct object-face relations, not involving a nodal stimulus, also will be similarly trained. In the scanner participants will be given experimental REP tests of SE in the form of CA (object-face) equivalence (transitivity plus symmetry) trials, together with the control REP tests of the direct object-face trained relations, in random order. The subtraction of the latter from the former should provide activations indicative of evocations of the nodal word stimuli as these are occurring during the emergence of equivalence.

Extensive training will enable a large number of SE and trained relations to be laid down and tested so that, within the initial sequence of tests and in each of a number of further independently randomized sequences, each specific relation will be tested once only. In this way the time course of any changes in the equivalence activations will be scrutinized. In terms of behavior, we predict progressive reductions in response latency for correct responses on successive trials of CA-same tests, and similarly on CA-different tests (though latencies on CA-different tests overall are predicted to be longer than those on CA-same tests). We theorize that such reductions may be due to two factors: (a) a facilitation in specific pathways due to repetition of tests of that specific pathway, perhaps due to short-circuits developing in the retrieval processes entailed; and (b) a facilitation of equivalence responses due to a reduction in the time required to switch in and execute the appropriate algorithm or relational frame of equivalence.

In terms of brain activity, we predict that because of factor (a), changes within a specific pathway as a function of repeated tests will lead to predictable attenuations in activations associated with activations of the nodal word stimulus, as reliance on evocation of the node becomes less essential to subsequent evocation of the face stimulus. Because of factor (b), deployment of the appropriate relational frame (RF) of equivalence will be facilitated by repetition, and we predict this will be associated with reduced activation in parietal clusters bilaterally, in DLPFC, especially on the left, and in the hippocampus.

CONCLUSION

The high technology of brain imaging and related techniques is both alluring and mystifying. It requires the behavior scientist to join forces with scientists from several other disciplines, including not only radiographers and physicists and neurophysiologists, but also those (neuro)psychologists who have been studying behavior from a cognitive perspective. The behavioral phenomena of derived relations seem to map on to the cognitive topic of inference, and may be close, particularly in the case of SE, to the fundamental transformations involved in language and symbolic processes. No explanation of behavior, at whatever level, will be valid in the long term, unless it accords with our best understanding of the activities of the brain. At the present time the possibility of such explanations is a far-away (to some, far-fetched) prospect, and it is daunting even to imagine what such explanations would be like. Studies like the ones described, and their prospective second-generation offspring, seem a sensible place from which to embark on this hazardous voyage of discovery.

REFERENCES

- Acuna, B. D., Eliassen, J. C., Donoghue, J. P., & Sanes, J. N. (2002). Frontal and parietal lobe activation during transitive inference in humans. *Cerebral Cortex*, *12*, 1312–1321.
- Acuna, B. D., Sanes, J. N., & Donoghue, J. P. (2002). Cognitive mechanisms of transitive inference. *Experimental Brain Research*, *146*, 1–10.
- Anderson, C., & Dickins, D. W. (2003, July). 'Nodal distance' versus 'symbolic distance': A contrast between two relational frames. Paper presented at the meeting of the European Association for Behavioural Analysis, Parma, Italy.
- Beason-Held, L. L., Purpura, K. P., Van Meter, J. W., Azari, N. P., Mangot, D. J., Optican, L. M., Mentis, M. J., Alexander, G. E., Grady, C. L., Horwitz, B., Rapoport, S. I., & Schapiro, M. B. (1998). PET reveals occipitotemporal pathway activation during elementary form perception in humans. *Visual Neuroscience*, *15*, 503–510.
- Bentall, R. P., Dickins, D. W., & Fox, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, *46B*, 187–214.
- Bentall, R. P., Jones, R. M., & Dickins, D. W. (1999). Control over emergent relations during the formation of equivalence classes: Response error and latency data for 5-member classes. *The Psychological Record*, *49*, 93–116.
- Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, *49*, 229–240.
- Bowlby, J. (1969). *Attachment and loss. Vol.1. Attachment*. New York: Basic Books.
- Bowlby, J. (1973). *Attachment and loss. Vol.2. Separation, anxiety, and anger*. New York: Basic Books.
- Buckmaster, C. A., Eichenbaum, H., Amaral, D. G., Suzuki, W. A., & Rapp, P. R. (2004). Entorhinal Cortex lesions disrupt the relational organization of memory in monkeys. *The Journal of Neuroscience*, *24*, 9811–9825.
- Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning*. New York: Wiley.
- Cerutti, D. T., & Rumbaugh, D. M. (1993). Stimulus relations in comparative primate perspective. *The Psychological Record*, *43*, 811–821.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic memory. *Psychological Review*, *82*, 407–428.
- Cullinan, V. A., Barnes-Holmes, D., & Smeets, P. M. (2001). A precursor to the relational evaluation procedure: Searching for the contextual cues that control equivalence responding. *Journal of the Experimental Analysis of Behavior*, *76*, 339–349.
- Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (7 May, 1999). Sources of mathematical thinking behavioral and brain-imaging evidence. *Science*, *284*, 970–974.
- Deutsch, C. K., Oross, S. P. I., DiFiore, A. W., & McIlvane, W. J. (2000). Measuring brain activity correlates of behavior: A methodological overview. *Experimental Analysis of Human Behavior Bulletin*, *18*, 36–42.
- Dickins, D. W. (2003, July). *The effects of a non-zero delay on response times in matching-to-sample (MTS) tasks in a stimulus equivalence (SE) paradigm*. Paper presented at the meeting of the European Association for Behaviour Analysis, Parma, Italy.
- Dickins, D. W., Bentall, R. P., & Smith, A. B. (1993). The role of individual stimulus names in the emergence of equivalence relations: The effects of paired-associates training between names. *The Psychological Record*, *43*, 713–724.
- Dickins, D. W., & Kramo, K. (2003). [Facilitatory effects of a short delay in delayed matching to sample: Evidence for a preparatory process in stimulus equivalence responding]. Unpublished raw data.
- Dickins, D. W., Singh, K. D., Roberts, N., Burns, P., Downes, J. J., Jimmieson, P., & Bentall, R. P. (2001). An fMRI study of stimulus equivalence. *Neuroreport*, *12*, 405–411.
- Dickins, D. W., Zaman, A., Bentall, R. P., Downes, J. J., Jimmieson, P., Mayes, A., & Roberts, N. (2002, June). *Brain activation during transitive inference in stimulus equivalence formation: a further fMRI study*. Paper presented at the International Conference on Functional Mapping of the Human Brain, Sendai, Japan.
- Dickins, T. E., & Dickins, D. W. (2001). Symbols, stimulus equivalence and the origins of language. *Behavior and Philosophy*, *29*, 221–244.
- Dickins, T. E., & Dickins, D. W. (2002). Is empirical imagination a constraint on adaptationist theory construction? *Behavioral and Brain Sciences*, *25*, 515–516.
- Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, *20*, 469–493.
- Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, *16*, 681–694.

- Eichenbaum, H., Otto, T., & Cohen, N. J. (1994). Two functional components of the hippocampal memory system. *Behavioral and Brain Sciences*, 17, 449–517.
- Fields, L., Reeve, K. F., Adams, B. J., & Verhave, T. (1991). Stimulus generalization and equivalence classes: A model for natural categories. *Journal of the Experimental Analysis of Behavior*, 55, 305–312.
- Fields, L., Reeve, K. F., Varelas, A., Rosen, D., & Belanich, J. (1997). Equivalence class formation using stimulus-pairing and yes-no responding. *The Psychological Record*, 47, 661–686.
- Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior*, 48, 317–332.
- Fletcher, P. C., & Henson, R. N. A. (2001). Frontal lobes and human memory: Insights from functional neuroimaging. *Brain*, 124, 849–881.
- Friston, K. J., Ashburner, J., Poline, J. B., Frith, C. D., Heather, J. D., & Frackowiak, R. S. J. (1995). Spatial registration and normalisation of images. *Human Brain Mapping*, 2, 165–189.
- Galizio, M., Stewart, K. L., & Pilgrim, C. (2001). Clustering in artificial categories: An equivalence analysis. *Psychonomic Bulletin & Review*, 8, 609–614.
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience*, 2, 568–573.
- Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, 118, 13–42.
- Grant, D. S., & Kelly, R. (2001). Anticipation and short-term retention in pigeons. In R. G. Cook (Ed.), *Avian visual cognition*: www.pigeon.psy.tufts.edu/avc/grant/.
- Greene, A. J., Spellman, B. A., Dusek, J. A., Eichenbaum, H. B., & Levy, W. B. (2001). Relational learning with and without awareness: Transitive inference using nonverbal stimuli in humans. *Memory & Cognition*, 29, 893–902.
- Greenham, S. L., Stelmack, R. M., & Campbell, K. B. (2000). Effects of attention and semantic relation on event-related potentials in a picture-word naming task. *Biological Psychology*, 55, 79–104.
- Halford, G. S., Bain, J. D., Maybery, M. T., & Andrews, G. (1998). Induction of relational schemas: Common processes in reasoning and complex learning. *Cognitive Psychology*, 35, 201–245.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001, September 28). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–2430.
- Haxby, J. V., Ungerleider, L. G., Horwitz, B., Rapoport, S. I., & Grady, C. L. (1995). Hemispheric differences in neural systems for face working memory: a PET-rCBF study. *Human Brain Mapping*, 3, 68–82.
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.) (2002). *Relational frame theory: A post-Skinnerian analysis of human language and cognition*. New York: Plenum.
- Hebb, D. O. (1958). Alice in wonderland or Psychology among the biological sciences. In H. F. Harlow, & C. N. Woolsey (Eds.), *Biological and Biochemical Bases of Behavior* (pp. 451–467). Madison, WI: University of Wisconsin Press.
- Heckers, S., Zalesak, M., Weiss, A. P., Ditman, T., & Titone, D. (2004). Hippocampal activation during transitive inference in humans. *Hippocampus*, 14, 153–162.
- Henson, R. N., Goshen-Gottstein, Y., Ganel, T., Otten, L. J., Quayle, A., & Rugg, M. D. (2003). Electrophysiological and haemodynamic correlates of face perception, recognition and priming. *Cerebral Cortex*, 13, 793–805.
- Henson, R. N. A. (2003). Neuroimaging studies of priming. *Progress in Neurobiology*, 70, 53–81.
- Holth, P., & Arntzen, E. (2000). Reaction times and the emergence of class consistent responding: A case for precurrent responding? *The Psychological Record*, 50, 305–337.
- Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior*, 65, 183–353.
- Horne, P. J., & Lowe, C. F. (1997). Toward a theory of verbal behavior. *Journal of the Experimental Analysis of Behavior*, 68, 271–296.
- Horwitz, B., Tagamets, M. A., & McIntosh, A. R. (1999). Neural modeling, functional brain imaging, and cognition. *Trends in Cognitive Sciences*, 3, 91–98.
- Huang, J., Carr, T. H., & Cao, Y. (2001). Comparing cortical activations for silent and overt speech using event-related fMRI. *Human Brain Mapping*, 15, 39–53.
- Imam, A. A. (2001). Speed contingencies, number of stimulus presentations, and the nodality effect in equivalence class formation. *Journal of the Experimental Analysis of Behavior*, 76, 265–288.
- Indefrey, P., & Levelt, W. J. M. (2000). The neural correlates of language production. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 845–865). Cambridge, MA: MIT Press.
- Ishai, A., Ungerleider, L. G., & Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. *Neuron*, 28, 979–990.
- Jahanshahi, M., & Rothwell, J. (2000). Transcranial magnetic stimulation studies of cognition: An emerging field. *Experimental Brain Research*, 131, 1–9.
- Josephs, O., & Henson, R. N. A. (1999). Event-related functional magnetic resonance imaging: Modelling, inference and optimization. *Philosophical Transactions of the Royal Society, London, Series B*, 354, 1215–1228.
- Josephs, O., Turner, R., & Friston, K. (1997). Event-related fMRI. *Human Brain Mapping*, 5, 243–248.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17, 4302–4311.
- Kastak, C. R., Schusterman, R. J., & Kastak, D. (2001). Equivalence classification by California sea lions using class-specific reinforcers. *Journal of the Experimental Analysis of Behavior*, 76, 131–158.
- Lazareva, O. F., Smirnova, A. A., Bagozkaja, M. S., Zorina, Z. A., Rayevsky, V. V., & Wasserman, E. (2004). Transitive responding in hooded crows requires linearly ordered stimuli. *Journal of the Experimental Analysis of Behavior*, 82, 1–19.
- Leader, G., Barnes-Holmes, D., & Smeets, P. M. (2000). Establishing equivalence relations using a respondent-type training procedure III. *The Psychological Record*, 50, 63–78.
- Lee, A. C. H., Robbins, T. W., Smith, S., Calvert, G. A., Tracey, I., Matthews, P., & Owen, A. M. (2002). Evidence for asymmetric frontal lobe involvement in

- episodic memory from functional magnetic resonance imaging and patients with unilateral frontal lobe excisions. *Neuropsychologia*, 40, 2420–2437.
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412, 150–157.
- Martin, N., & Alsop, B. (2004). Transitive inference and awareness in humans. *Behavioural Processes*, 67, 157–165.
- Mayes, A. R., Gooding, P. A., & van Eijk, R. (1997, June). A new theoretical framework for explicit and implicit memory. *PSYCHE*, 3(2). Retrieved from <http://psyche.cs.monash.edu.au/v3/psyche-3-02-mayes.html>
- McGonigle, B., & Chalmers, M. (1992). Monkeys are rational! *Quarterly Journal of Experimental Psychology: B*, 45, 189–228.
- Menon, R. S. (2001). Imaging function in the working brain with fMRI. *Current Opinion in Neurobiology*, 11, 630–636.
- Menon, V., Adelman, N. E., White, C. D., Glover, G. H., & Reiss, A. L. (2001). Error-related brain activation during a go/nogo response inhibition task. *Human Brain Mapping*, 12, 131–143.
- Menon, V., Mackenzie, K., Rivera, S. M., & Reiss, A. L. (2002). Prefrontal cortex involvement in processing incorrect arithmetic equations: evidence from event-related fMRI. *Human Brain Mapping*, 16, 119–130.
- Nagode, J. C., & Pardo, J. V. (2002). Human hippocampal activation during transitive inference. *Neuroreport*, 13, 939–944.
- Nixon, P., Lazarova, J., Hodinott-Hill, I., Gough, P., & Passingham, R. (2004). The inferior frontal gyrus and phonological processing: An investigation using rTMS. *Journal of Cognitive Neuroscience*, 16, 289–300.
- O'Donnell, J., & Saunders, K. J. (2003). Equivalence relations in individuals with language limitations and mental retardation. *Journal of the Experimental Analysis of Behavior*, 80, 131–157.
- Paz-Y-Mino, C. G., Bond, A. B., Kamil, A. C., & Balda, R. P. (2004). Pinyon jays use transitive inference to predict social dominance. *Nature*, 430, 778–781.
- Place, U. T. (1995/1996). Symbolic processes and stimulus equivalence. *Behavior and Philosophy*, 23/24, 13–30.
- Posner, M. I., & Mitchell, R. F. (1967). Chronometric analysis of classification. *Psychological Review*, 74, 392–409.
- Postle, B. R., & D'Esposito, M. (2000). Evaluating models of the topographical organisation of working memory function in frontal cortex with event-related fMRI. *Psychobiology*, 28, 132–145.
- Price, C. J., Wise, R. J., & Frackowiak, R. S. (1996). Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral Cortex*, 6, 62–70.
- Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2003). Dissociating Explicit and Implicit Category Knowledge with fMRI. *Journal of Cognitive Neuroscience*, 15, 574–583.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
- Rissman, J., Eliassen, J. C., & Blumstein, S. E. (2003). An event-related fMRI investigation of implicit semantic priming. *Journal of Cognitive Neuroscience*, 15, 1160–1175.
- Rumsey, J. M., Horwitz, B., Donohue, B. C., Nace, K., Maisog, J. M., & Andreason, P. (1997). Phonologic and orthographic components of word recognition: A PET-rCBF study. *Brain*, 120, 739–759.
- Rushworth, M. F. S., & Owen, A. M. (1998). The functional organization of the lateral frontal cortex: Conjecture or conjuncture in the electrophysiology literature? *Trends in Cognitive Sciences*, 2, 46–53.
- Schusterman, R. J., & Kastak, D. (1993). A California sea lion (*Zalophus californianus*) is capable of forming equivalence relations. *The Psychological Record*, 43, 823–839.
- Schusterman, R. J., & Kastak, D. (1998). Functional equivalence in a California sea lion: Relevance to animal social and communicative interactions. *Animal Behaviour*, 55, 1087–1095.
- Sevostianov, A., Hortwitz, B., Nechaev, V., Williams, R., Fromm, S., & Braun, A. R. (2002). fMRI study comparing names versus pictures of objects. *Human Brain Mapping*, 16, 168–175.
- Shallice, T. (2003). Functional imaging and neuropsychology findings: How can they be linked? *NeuroImage*, 20, S146–S154.
- Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston, MA: Authors' Cooperative.
- Sidman, M., & Tailby, W. (1982). Conditional discrimination vs matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37, 5–52.
- Skinner, B. F. (1989). The origin of cognitive thought. *American Psychologist*, 44, 13–18.
- Smith, A. B., Dickins, D. W., & Bentall, R. P. (1996). The role of individual stimulus names in the emergence of equivalence relations II: The effects of interfering tasks prior to and after tests for emergent relations. *The Psychological Record*, 46, 109–130.
- Smith, E. E., & Jonides, J. (12 March, 1999). Storage and executive processes in the frontal lobes. *Science*, 283, 643–659.
- Spencer, T. J., & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior*, 65, 643–659.
- Staunton, C., Barnes-Holmes, D., Whelan, R., & Barnes-Holmes, Y. (2003, May). *Priming and event related potentials (ERPs) as measures of derived relational responding*. Paper presented at the Association for Behavior Analysis 29th Annual Convention, San Francisco, CA.
- Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). New York: Wiley.
- Stich, S. P. (1983). *From folk psychology to cognitive science: The case against belief*. Cambridge, MA: MIT Press.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: An approach to cerebral imaging*. (M. Rayport, Trans.). New York: Thieme Medical.
- Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20, 410–433.
- Treichler, F. R., Raghanti, M. A., & Van Tilburg, D. N. (2003). Linking of serially ordered lists by macaque monkeys (*Macaca mulatta*): List position influences. *Journal of Experimental Psychology: Animal Behavior Processes*, 29, 211–221.

- Von Fersen, L., Wynne, C. D. L., Delius, J. D., & Staddon, J. E. R. (1991). Transitive inference formation in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 17, 334–341.
- Walsh, V., & Pascual-Leone, A. (2003). *Transcranial magnetic stimulation: A neurochronometrics of mind*. Cambridge, MA: MIT Press.
- Warren, A. M., & McIlvane, W. J. (1998, April). *Stimulus equivalence and the N400 effect*. Paper presented at the meeting of the Cognitive Neuroscience Society, San Francisco, CA.
- Wright, A. (2001). Learning strategies in matching to sample. In R. G. Cook (Ed.), *Avian Visual Cognition*. www.pigeon.psy.tufts.edu/avc/wright.
- Wright, B. C. (2001). Reconceptualizing the transitive inference ability: A framework for existing and future research. *Developmental Review*, 21, 375–422.
- Wynne, C. D. L. (1995). Reinforcement accounts for transitive inference performance. *Animal Learning and Behavior*, 23, 207–217.

Received August 27, 2004

Final Acceptance April 23, 2005