Behavior and Philosophy, Winter 1995/1996, Volume 23 No.3/Volume 24 No. 1, pp. 13-30

SYMBOLIC PROCESSES AND STIMULUS EQUIVALENCE[1]

Ullin T. Place
University of Wales Bangor

ABSTRACT: A symbol is defined as a species of sign. The concept of a sign coincides with Skinner's (1938) concept of a discriminative stimulus. Symbols differ from other signs in five respects:
(1) They are stimuli which the organism can both respond to and produce, either as a self-directed stimulus (as in thinking) or as a stimulus for another individual with a predictably similar response from the recipient in each case.
(2) they act as discriminative stimuli for the same kind of object for all members of the verbal community within which they function as symbols;
(3) they acquire their properties by virtue of arbitrary social convention rather than any natural and intrinsic connection between the sign and what it is a sign of;
(4) competent members of the verbal community can both produce the appropriate symbol in response to a naturally occurring sign of the presence of the object or a sample of the kind of object which the symbol stands for and select the appropriate object when presented with the symbol;
(5) they form stimulus equivalence classes of the kind demonstrated in the matching-to-sample task (Sidman, 1971; Sidman and Tailby, 1982) both with other symbols having the same meaning and, more important, with the naturally-occurring non-symbolic signs of the presence of the object or kind of object which the symbol stands for.

Connectionism and the Externalist Theory of Symbolic Processes

In a previous article in this journal (Place, 1992) I argued that the connectionist revolution in artificial intelligence should lead to a return to the empiricist-behaviorist linguistics that flourished before Chomsky (1959), in his well known review of Skinner's (1957) book *Verbal Behavior*, persuaded linguists and other students of language that the empiricist-behaviorist approach could not hope to succeed. The grounds for this prediction come from a comparison between the digital or von Neumann computer which is the model underlying Chomsky's theory and the connectionist network that has increasingly replaced it as the preferred model for the way the brain operates. A digital computer is a symbolic processor. It generates an output by converting the strings of symbols fed into it by a human operator into a digital code, the so-called "machine language", and computing the

---

[1] AUTHOR'S NOTE
Earlier versions of this paper were presented at the Annual Conference of the Experimental Analysis of Behaviour Group, Department of Psychology, University College, London, 5th April 1991, the 17th Annual Convention of the Association for Behavior Analysis, Atlanta, Georgia, May 26th 1991, and the 15th Symposium on Quantitative Analyses of Behavior on 'Stimulus Relations', Harvard University, Cambridge, MA, June 14th 1992. Requests for offprints to U.T. Place, Willowtree Cottage, Boltby, Thirsk, North Yorkshire, YO7 2DY, England.

output by conforming strictly to a set of symbolically formulated rules which constitute the operative computer program. When used as a model for the generation and interpretation of language by the brain, it favors an *internalist* theory of the symbols and symbolic processing which characterize the forms of thought and communication that are typically and exclusively human. According to this theory symbols, their manipulation and the rules governing those manipulations are construed as part of the internal furniture of the brain just as they are part of the internal furniture of the device which the theory takes as a model.

In a connectionist network, on the other hand, symbols and symbolic processing play no part. An output is generated from an input which is nothing more than a pattern of sensory excitation by the way the input is transformed as it passes through the network of synaptically interconnected 'nodes', the analogue of the neurons that make up the 'grey matter' of the nervous system, as determined by the so-called "weights" of the individual connections. When used as a model for the generation and interpretation of language by the brain, it favors an *externalist* theory of the symbols and symbolic processing which characterize the forms of thought and communication that are typically and exclusively human. According to this theory the primary location of symbols, their manipulation and the rules governing those manipulations is in the environment. Symbols are construed as responses emitted by one organism, the speaker or writer, which act as stimuli controlling the behavior of another organism, the listener or reader. The process of symbol manipulation or thinking is construed as a case where the speaker directs her symbol emission response towards herself and thus combines in one person the roles of speaker and listener. Such self-directed speech typically occurs as a private internalized process; but thinkers can and often do talk to themselves aloud, particularly in the early stages of language acquisition. The rules which govern such manipulations to which speakers and thinkers must conform if the results of such manipulations are to be effective in controlling both their physical and their social environment are to be construed as social conventions. Such conventions are learned in the first place and subsequently maintained by the reinforcement and error-correcting practices of the verbal community in which they are embodied.

It goes without saying that this externalist theory of symbols and symbolic processing coincides exactly with the traditional empiricist-behaviorist theory of language which was abandoned by mainstream linguistics in favor of Chomskean internalism more than thirty five years ago. But if I am right in thinking that the connectionist revolution in artificial intelligence points in the direction of a return to an externalist theory of symbols and symbolic processing and an empiricist-behaviorist linguistics, we cannot expect simply to return to the situation as it was forty years ago. Although we can reasonably expect the internalist paradigm to be increasingly discarded, it must be conceded that much of permanent value has been achieved over the past thirty five years by those working within it. In particular, as I argued in the previous article and as Skinner (1989) was beginning to concede in one of the last papers he published before he died, the behaviorist must accept that Chomsky is right when he claims that the functional unit of language is the sentence and that sentences are seldom repeated word for word, but are typically constructed anew on each occasion of utterance.

The advances that have been made over the past thirty five years have not all come from within the internalist paradigm. There are at least two important contributions that have come from within the behavior analytic tradition. One of these is the distinction which Skinner (1966/1969/1988) draws between "contingency-shaped" and "rule-governed" forms of behavior in his 'An operant analysis of problem solving.' Another is the work of Murray Sidman (1971; 1986; 1990; 1992), his collaborators and those who have followed his lead on the phenomenon of "stimulus equivalence" as observed in the matching-to-sample experiment. It is with this latter development and its implications for an externalist theory of symbols and symbolic processing that this article is concerned.

Signs and Symbols as Discriminative Stimuli

Any theory of symbolic communication and symbolic thinking must begin by examining the concept of symbol itself. What, then, is a symbol? A symbol is a species of sign. All symbols are signs, but not all signs are symbols. But what, then, is a sign? A sign involves three things:

(1) the *sign* itself, a behavior-controlling change in the stimulus environment of

(2) an *organism* for which it is a sign, and

(3) a *contingency*, a causal relation between behaving in a certain way on the one hand and the consequences of so behaving on the other to whose existence the behavior of the organism is thereby alerted or orientated.

The concept of a sign so defined coincides closely with Skinner's (1938) concept of a discriminative stimulus. The only important difference between the two is that Skinner's concept is tied, in the way the concept of a sign is not, to a particular discrimination learning procedure whereby stimuli acquire their properties as S^D-s or S^Δ-s with respect to the emission and reinforcement or non-reinforcement of a particular response. It now turns out that this discrimination learning procedure is formally identical with the process of supervised learning or learning by error-correction whereby a neural network or PDP learns to recognize and respond to different patterns of input.[1] The remarkable convergence of these two research traditions makes it tolerably certain that operant *alias* supervised discrimination learning is the basic learning procedure whereby signs in general and symbols *qua* species of sign acquire their behavior-orientating function.

The Problem of Novel Sentence Construction

So long as it remains tied to the process of operant discrimination learning, the concept of a discriminative stimulus cannot begin to handle the phenomenon to which Chomsky (1957, etc.) has repeatedly drawn attention, the phenomenon of the *novel sentence*. This is the phenomenon whereby, provided certain grammatical (or, to use Skinner's 1957 term, "autoclitic") conventions are adhered to, a string of words which the listener has never encountered before can, nevertheless, act as a sign for the existence of a contingency the like of which he or she need never have encountered. To the extent that it orientates the behavior of the organism towards the impending presence of some contingency, a sentence acts as a sign; but in so far as the listener has never encountered that particularly combination of words before, the behavior-orientating property of the sign cannot be explained by repeated pairing of the sign and what it stands for. Discrimination learning can be invoked to explain the way sentences act as signs for the contingencies they depict only to the extent that it explains the contribution that is made to the meaning of the sentence as a whole by those constituents of sentences which *are* repeated. The sentence constituents that are repeated are the noun and verb phrases used to identify the objects and events involved in the contingency depicted by the sentence as a whole, and the syntactic (autoclitic) structure of the sentence which depicts the way those objects and events are related.

Words sometimes acquire their sign-function or meaning by means of definitions constructed of other words. But not all words can acquire their meaning in this way; for you can learn the meaning of a word from a definition, only if you already understand the meaning of the words of which the definition is composed. There is, therefore, no alternative to the view that words which make up a child's basic vocabulary acquire their meaning in the first instance by the same process of operant or supervised discrimination learning as is encountered in the behavior of animals and neural networks.

Words are a species of symbol. All words are symbols, but not all symbols are words. All symbols, however, can be said to have a conventional meaning that is shared by a word-using human community within which the same natural language is spoken. It may well be the case that some symbols that are not words, road signs for example, acquire their meaning directly through a process of discrimination learning, such as that in which the matching-to-sample task consists. Nevertheless, there would appear to be no symbols whose meaning *cannot* be expressed in words. This is part of the evidence for the view that all symbols and symbolic functions either consist in or involve the pre-existence of natural language as system of interpersonal communication.

Symbols as a Species of Sign

What is it that marks off words and other symbols from other discriminative stimuli or signs? In order to answer that question we need to consider what it is that signs and symbols have in common and that justify us in regarding symbols as a species of sign.
We have seen that symbols are like signs in two respects:

(1) they have *meaning* or intentionality, to use a technical philosophical term that is often invoked in this connection, in that *qua* stimuli they orientate the behavior of the organism towards the impending presence or existence of some actual or possible feature of the environment, and
(2) they acquire that meaning either directly or indirectly through some process of *discrimination learning*.

Symbols differ from other signs in that

(1) whereas non-symbolic signs are signs only for those individuals who have been exposed to the relevant discrimination learning experiences, symbols have a meaning that is
    (a) the same for all fully-fledged members of the *social group* (the verbal community) within which it functions as a symbol,
    (b) *not limited*, as are non-symbolic signs, to those features of the world which the individual has experienced,
    (c) fixed and maintained by the *linguistic conventions* and consequent error-correcting practices that prevail within the verbal community, and
    (d) connected to its referent by *arbitrary social convention*,
(2) whereas non-symbolic signs are either accidental concomitants of the objects, events or states of affairs of which they are signs, like someone's shadow or their footfall, or, if non-accidental, like the secondary sexual characteristics which attract one sex to another, are adapted to that function as a consequence of the effects of natural selection on the genetic constitution of both sign producer and sign recipient, symbols are either *human artifacts* specifically constructed for that purpose or, where they are naturally-occurring objects such as human individuals, mountains, rivers or springs,

acquire their symbolic function through their place in a system of symbolic artifacts and ritual social practices, and

(3) whereas a non-symbolic sign, such as someone's visual appearance or the sound of their voice, has a direct, non-arbitrary, asymmetrical and intransitive connection to the kind of object, event or state of affairs of which it is the sign, symbols are connected to the things they symbolize *indirectly* through an arbitrary, symmetrical and transitive *equivalence relation* between the symbol and one or more non-symbolic and non-arbitrary signs of the kind of object, event or state of affairs it symbolizes.

It seems that the process whereby a language learning organism (a human child) learns to treat an arbitrary symbol as equivalent to a naturally-occurring non-symbolic sign (or 'natural sign', as I shall say henceforward) of the presence of some object, event and state of affairs is the same process as that whereby *stimulus equivalence classes* are formed in the matching-to-sample experiment as described by Sidman (Sidman, 1971; Sidman and Tailby, 1982; Sidman, 1986; 1990; 1992).

Symbols and the Formation of Stimulus Equivalence Classes

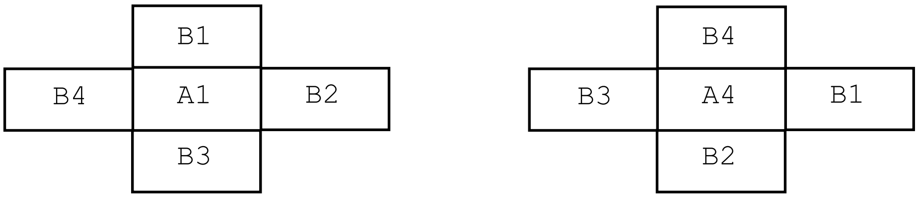


Figure 1. First Training Session-Trials 1 & 4.

Figures 1-5 are designed to assist those unfamiliar with the matching-to-sample experiment and with the concept of a *stimulus equivalence class* as defined in terms of responding in that experimental situation. The boxes labelled A1-A4, B1-B4 and C1-C4 represent arbitrarily selected visually presented shapes. These appear on a screen in the subjects line of regard in groups of five, shown by the four crosses on each figure. These crosses two of which appear on each figure represent two of the four trial types (numbers 1 and 4) into which each training session or test is divided. The letter and number combinations A1-A4, B1-B4 and C1-C4 are simply for purposes of identification and do not correspond to any sequential order perceptible to the subject. The stimulus in the center of the display is illuminated before the other four and is known as the *sample*. The four peripheral shapes (the precise number is not significant) appear simultaneously shortly after the exposure of the sample and are known as the *comparison stimuli*. The subject's task is to pick from the comparison stimuli the shape which, by arbitrary *fiat* on the part of the experimenter, is designated as the correct choice. In Figures 1-5 the correct choice is the one shown in the box at the top of each cross. In practice, however, the position of the four comparison stimuli is changed from trial to trial in a random order so as to prevent responding on the basis of position rather than the character of the stimulus selected. The subject's task is discover which response is correct from the occurrence or non-occurrence of a correct message or reinforcement which is received when and only when the correct choice is made. In the case of an adult or older child this correct message may consist in hearing the word *Right!* or *Correct!* or in the case of a younger child in the occurrence of an exciting and attractive audio-visual event. An incorrect choice will be indicated either by the absence of the reinforcer or by a positive disinforcer such as the word *Wrong!*

After a few training trials with this regime, incorrect responses will be eliminated and the subject will select the correct shape (B1-4) every time. Once this pattern of responding is established, two test trials (Figures 2 and 3) are interposed. The purpose of these tests is to determine whether the subject has learned only what an animal would learn in these circumstances, namely to pick a B when presented with an A as sample, or whether he or she has learned to do what a human adult tends to do when confronted by such an experience, namely to treat the corresponding A's and B's as equivalent to one another for the purposes of the experimental situation.

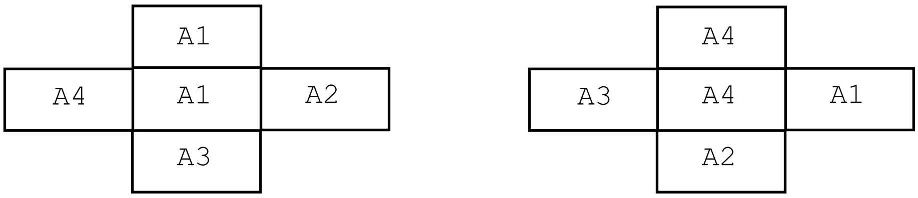


Figure 2. Reflexivity Test-Trials 1 & 4.

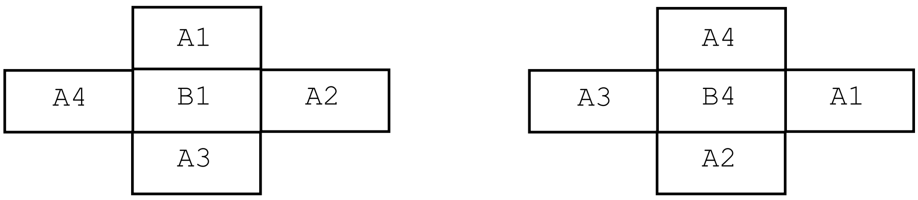


Figure 3. Symmetry Test-Trials 1 & 4.

In logic, the equivalence relation is characterized by three features:

(1) *reflexivity* (For the same value of A, A is equivalent to A),

(2) *symmetry* (If A is equivalent to B, B is equivalent to A), and

(3) *transitivity* (If A is equivalent to B, and B is equivalent to C, A is equivalent to C).

Hence, if as a consequence of having been taught to pick a B when presented with an A as sample, the subject has come to treat the A as equivalent to the corresponding B, it should follow

(a) that when presented with the *reflexivity test* (Figure 2), with an A as sample he or she should spontaneously pick the A on the first presentation without any prompting or specific training so to do,

(b) that when presented with *symmetry test* (Figure 3), with a B as sample he or she should likewise spontaneously pick A on the first presentation without any prompting or specific training so to do. After a further training session (Figure 4) in which the subject learns to pick a C, when a B is presented as sample, it is predicted

(c) that when presented with the *transitivity test* (Figure 5), with an A as sample he or she should spontaneously pick C on the first presentation without any prompting or specific training so to do.

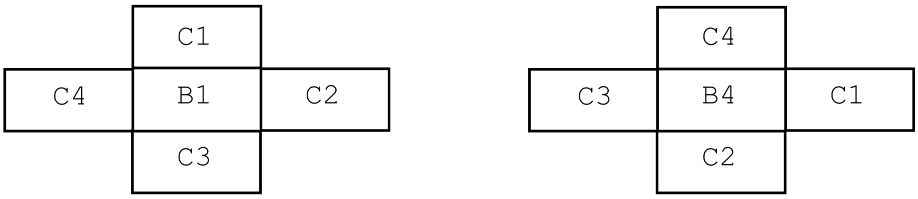


Figure 4. Second Training Session-Trials 1 & 4.

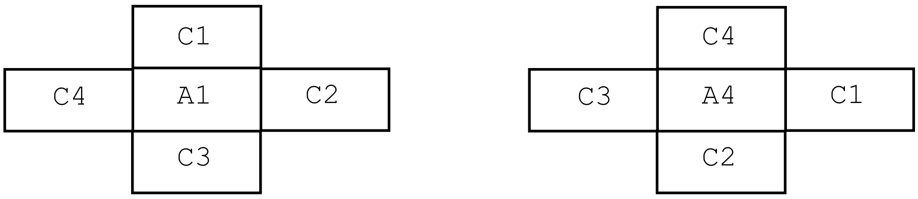


Figure 5. Transitivity Test-Trials 1 & 4.

Inspired by evidence that children begin to form such stimulus equivalence classes at a very early stage of linguistic development, while animals in general never do, the belief that the spontaneous formation of stimulus equivalence classes on the matching-to-sample task is intimately bound up with the process whereby arbitrary stimuli acquire the property of acting as symbols has been part and parcel of this experimental research tradition since its inception. Sidman's original idea, influenced, unconsciously perhaps, by the serial-digital computer model of brain functioning which pervaded psychology at the time, was that the development of the child's ability to form arbitrarily selected stimuli into equivalence classes preceded and made possible the subsequent acquisition of linguistic competence. Subsequent research (Lowe and Beasty, 1987; Beasty, 1987; Dugdale and Lowe, 1990, Horne and Lowe, forthcoming), however, has shown beyond all reasonable doubt that this is not the case. Some degree of linguistic competence must already be present, before equivalence classes begin to be formed on the matching-to-sample task. Nevertheless, the idea that, in exploring the formation of equivalence classes, we are exploring the fundamental process whereby arbitrary symbols acquire their meaning is deeply rooted, not only in Sidman's own thinking, but in that of those who have followed his lead into this exciting research area.

Equivalence Classes in Logic and Behavior Analysis

I confess that, until comparatively recently, I had deep misgivings about this idea. This was because the relation between a symbol and what that symbol symbolizes is *not* an equivalence relation. Take, for example, a proper name such as the name "Margaret Thatcher." If we apply the standard tests of equivalence, reflexivity, symmetry and transitivity, we find that this relation fails all of them. The name "Margaret Thatcher" does not stand for itself. The person Margaret Thatcher does not stand for the name "Margaret Thatcher" in the way that the name stands for its bearer. From the fact that "Margaret Thatcher" is the married name of Margaret Roberts and "Margaret Roberts" is the name of Arthur Roberts' wife, it does not follow that "Margaret Thatcher" is the name of Arthur Roberts' wife.

However, in the light of a discussion with Lanny Fields when we first met at a meeting of the Experimental Analysis of Behaviour Group at Cambridge, England in April 1989, I now accept that, although a symbol does not form an equivalence class with the object, event or state of affairs that it symbolizes, it *does* form an equivalence class with the naturally occurring non-symbolic signs of the presence of that object, event or state of affairs. Thus although, as we have seen, the name "Margaret Thatcher" does not form an equivalence class with the bearer of that name, it does form an equivalence class with the naturally occurring non-symbolic signs of the presence of its bearer such as her visual appearance and the sound of her voice.

This insight goes some way towards meeting the misgivings which I had about the significance of the concept of stimulus equivalence in relation to the concepts of 'a symbol' and 'symbolic function'. But there is still a problem. For from the standpoint of logic (which is, after all, the discipline from which the concepts of 'equivalence relation' and 'equivalence class' derive) what makes a person's name part of the same equivalence class as her visual appearance and the sound of her voice is simply that they all function as discriminative stimuli for an encounter with the same person. In this respect they are no different from the visual appearance, sound and smell of, say, a leopard. These too all have the same effect, that of alerting members of a species for whom the leopard is a dangerous predator to the presence of such a creature in the vicinity. Here, the visual appearance, sound and smell characteristic of a leopard form an equivalence class simply by virtue of the fact that they all evoke the *same* response. All the tests of equivalence are satisfied. Reflexivity is satisfied by the fact that different tokens of, say, the visual appearance typical of leopards all evoke the same response. Symmetry is satisfied by the fact that if a token of the visual appearance of the leopard type orientates behavior in the same way as does any token of the typical leopard sound, we can infer that any token of the typical leopard sound will orientate behavior in the same way as does any token of the visual appearance typical of leopards. Transitivity is satisfied by the fact that if a particular visual appearance A orientates behavior in the same way as does a sound B, and sound B orientates behavior in the same way as does smell C, we can infer that visual appearance A orientates behavior in the same way as does smell C. We can substitute the name "Margaret Thatcher," Margaret Thatcher's visual appearance and the sound of her voice for the sight, sound and smell typical of leopards in this example without affecting the validity of the argument in any way.

There is, however, an important difference between the stimulus equivalence class constituted by the sight, sound and smell of a leopard and the stimulus equivalence class constituted by the name "Margaret Thatcher" and her visual appearance and the sound of her voice. The latter, since it includes an arbitrary symbol in the shape of the name "Margaret Thatcher" qualifies as a stimulus equivalence class in Sidman's sense, whereas the former, since it includes only natural signs, does not. The sight, sound and smell of a leopard are equivalent only in the sense that they all evoke the same behavior, behavior appropriate to escaping from leopards. The relations of reflexivity, symmetry and transitivity which define an equivalence class are manifested in this case only in the inferences that *we* draw in arguing from the effect of one stimulus to the effect of another. For the organism, the only link between the members of the class is that they are all signs for the same contingency.

The difference between an equivalence class in the logical sense to which stimuli belong by virtue of having the same effect on the behavior of an organism and a stimulus equivalence class in Sidman's sense is not easy to define. What is clear is that learning to respond in the same way, e.g., picking a member of the stimulus class B1-B4 whenever a member of the stimulus class A1-A4 is present, is not what is meant. That is just the training procedure which results in the *formation* of a Sidman stimulus equivalence class. But it only does so in an organism that is capable of forming such equivalence classes; and at present we have indisputable evidence for such equivalence class formation only in the case of human subjects who have attained a certain level of linguistic competence. Moreover, it is an essential feature of a Sidman stimulus equivalence class that its membership *extends beyond the stimuli to which the subject has been trained to respond* (the samples) to include those which she has been trained to select (the comparisons). When the equivalence class is formed, the stimuli used in the training sessions become interchangeable with respect to their two functions:

(a) that of a discriminative stimulus *controlling* the response, and

(b) that of a target stimulus the selection or 'picking' of which *constitutes* the response.

The significance of this interchangeability for the theory of symbolic functioning derives from the observation that outside the laboratory it is a property that applies only between a symbol and its equivalent in another sensory modality, e.g., the spoken and written versions of the word "dog," between a symbol and its equivalents in other languages, e.g., "chien," "hund," "perro," and between all these and the natural signs of the presence of an instance of the kind in question, e.g., the characteristic visual appearance presented by dogs and the typical barking sound they make. In all these cases a subject who has formed the equivalence class can select one member of the class when presented with another regardless of which is presented as sample and which as comparison.

Arbitrary Symbols and Natural Signs

Exactly why we should find this interchangeability between the functions of discriminative stimulus and selection target applies only in the case of symbols, their equivalents and the natural signs of the presence of their referents is not altogether clear. If we take the matching-to-sample experiment as a model, it suggests that it is connected with the arbitrary character of the symbol which contrasts with the non-arbitrary character of the connection between the natural signs and the objects and situations to whose presence both kinds of stimuli alert the responding organism. Although this does not appear from the matching-to-sample experiment, the importance of a symbol's being arbitrary is connected to the ease with which readily discriminable versions of such stimuli can be generated by a speaker, writer or thinker whenever she needs to draw the listener's or her own attention to the symbol's referent. This contrasts with the fact that in the days before the development of paper on which to draw a rapid sketch, photography and video-recording the only natural signs that could be readily reproduced by human communicators were auditory stimuli such as those generated by imitating the calls of birds or animals or such things as the sound of water being poured from a bottle.

But what is also clear from the matching-to-sample experiment is that the stimulus equivalence classes between arbitrarily selected stimuli which are formed in such experiments are of no functional significance for the subject outside that artificial situation. What this tells us is that a Sidman stimulus equivalence class acquires genuine meaning and its arbitrarily linked members begin to function as genuine symbols only when the equivalence class includes either

(a) a natural sign of the presence of some individual or instance of a kind for the presence of which the arbitrarily connected members of the class thereby act as a symbol, or
(b) a word which possesses an existing symbolic meaning by virtue of the conventions of the natural language to which it belongs and is recognized as such by the subject.

In other words, a word possesses the conventional symbolic meaning it does, if and only if it, or some other word in terms of which its meaning is defined or otherwise explained, is a member of a Sidman stimulus equivalence class which contains one or more natural signs of the presence of the individual or kind which the sign thereby comes to designate.

Generalizing from this discussion, we can state what we may call

The Fields Principle

> A symbol is a sign which designates its object, not by virtue of a naturally-occurring association between sign and *significatum*, but by virtue of its membership of a Sidman stimulus equivalence class among whose other members is at least one sign which *does* derive its semantic function from such a naturally-occurring association and transfers that function to other members of the class.

Stimulus Equivalence and Naming

Important as it is for our understanding of the relation between symbols and the formation of equivalences classes on the matching-to-sample task, all that the Fields Principle tells us is that in natural circumstances outside the laboratory an arbitrary sign will acquire meaning only by virtue of establishing a link with some other sign whose connection with the contingency to whose presence the sign alerts the organism is not arbitrary. What the Fields Principle does not tell us is

(a) why symbols, as arbitrary signs, should be linked to the natural signs that give them their meaning in this special way such that the linkage generalizes spontaneously in the manner which Sidman describes as the formation of a stimulus equivalence class,
(b) why there should be this apparently invariable association in nature between responding to and producing arbitrary signs whose meaning is assigned by social convention (symbols) and the formation of those symbols into Sidman stimulus equivalence classes,
(c) why these phenomena should be restricted, as they seem to be, to human beings, and
(d) why the association between stimulus equivalence and the use of symbols should develop, as it seems to do, at a relatively early stage in the child's acquisition of linguistic competence?

At the present time there are three schools of thought on these issues. One is the view of Sidman (1986) and his colleagues at the E. K. Shriver Center in Massachusetts. This is the view that the ability to form stimulus equivalence classes is an innate ability peculiar to human beings which precedes the acquisition of the ability to produce intelligible speech and makes it possible. Another is the Relational Frame Theory advocated by Steve and Linda Hayes (Hayes, 1991; Hayes and Hayes, 1992) and their students at the University of Nevada, Reno. This agrees with the Sidman position in holding that the ability to form stimulus equivalence classes precedes the acquisition of language and makes it possible though it identifies the two processes rather more closely than does Sidman. It differs from the earlier view principally in regarding equivalence as only one among a number of relations between stimulus events which the child learns to identify and respond to before and in the course of acquiring language. The third position and the one advocated in this paper is the view developed by Fergus Lowe and his colleagues at the University of Wales Bangor.

This work takes as its starting point a principle whose reality was first demonstrated at Bangor under Professor Lowe's supervision by Alan Beasty (Lowe and Beasty, 1987; Beasty, 1987):

The Principle of the Mediation of Stimulus Equivalence by Naming

> The spontaneous emergence of equivalence responding on the matching-to-sample task in children who are in the early stages of acquiring linguistic competence depends on their learning a distinguishing name for each of the stimuli employed in the task.

A second principle which emerged from subsequent discussions at Bangor between Fergus Lowe, Neil Dugdale, Pauline Horne and Julie Hird draws attention to a respect in which names - and, I would add, symbols in general - differ from ordinary non-symbolic discriminative stimuli or signs (Dugdale and Lowe, 1990):

The Principle of Reciprocal Name-Object Selection

> When a child learns a name for an object, it learns not only to produce the name when presented with the object, it also learns to pick out the object when presented with the name, and to treat both as parts of a single task, that of connecting the object with its name.

What this principle shows is that in learning the names of things, the child is learning to do something that a pre-linguistic organism does not normally learn to do.[2] It learns to respond symmetrically to two stimuli, an arbitrary stimulus, the name, and a natural sign of the presence of the object it is the name of, the object's visual or auditory appearance. It learns to produce or pick out the one in the presence of the other and *vice versa*. In other words, in learning the names of things, the child learns to make the kind of response that would be judged correct, if made by the subject in the matching-to-sample experiment when tested for symmetry (Figure 3 above), and would thus constitute a substantial part of the evidence for the subject's having formed a Sidman stimulus equivalence class.

Moreover, if, as happened in the Beasty experiment, it is trained to name the stimuli which it subsequently forms into an equivalence class, the child would have received just the kind of pre-training which, if given to an animal, would be judged to disqualify the claim that, on the basis of nothing more than the first training session (Figure 1 - given A as sample, pick B), it had *spontaneously* selected the correct response on the symmetry test (given B as sample, pick A) on the first presentation.

This evidence must raise the suspicion, at least in the case of children who are in the early stages of linguistic development, that passing the various tests for the formation of a Sidman stimulus equivalence class may not in fact represent anything like a grasp on the part of the child of the notion that the stimuli in question are to be treated as equivalent. It may simply be that the kind of training the child receives when learning to name objects in general and the test stimuli, in particular, pre-disposes it respond in the test situations in the same way as it would be expected to do if it *had* acquired that concept. It is true that this suspicion arises in the first instance in relation to only one of the three tests of equivalence, the symmetry test. A more detailed analysis, however, suggests that learning to name the stimuli involved may in fact mediate what we should perhaps call `pseudo-equivalence responding' on all three tests. This pseudo-equivalence responding would then contrast with the kind of responding on such tests which is typically, but by no means always (Cf. Adams, Fields and Verhave, 1993; Fields, Adams and Verhave, 1993), found in the responding of older children and adults on these tests which is undoubtedly based on grasping the idea that the stimuli involved are to be treated as equivalent and hence interchangeable with one another

The Role of Naming in Pseudo-Equivalence Responding

In order to explain how a child's learning to name the stimuli involved might generate pseudo-equivalence responding on all three test of stimulus equivalence we need to invoke two principles:
(A) *The Principle of Symmetrical Responding*; and
(B) *The Principle of Anticipatory Naming*.

These principles can be stated as follows:

The Principle of Symmetrical Responding

> If the response of picking a stimulus of type B is regularly reinforced in the presence of a sample stimulus of type A and other responses are not reinforced under that condition, and if, moreover, all the stimuli involved in the discrimination learning process evoke an utterance of a name distinctive of stimuli of that type, the subject will spontaneously and without previous reinforcement of that response pick a stimulus of type A in the presence of a stimulus of type B which it would not do if the stimuli had not been named.[3]

The Principle of Anticipatory Naming

> If the response of picking B is regularly reinforced in the presence of A, and if other responses are not reinforced under that condition, and if both stimuli are given a distinctive name by the subject, the occurrence of A, in addition to evoking an utterance of its name, will also evoke an utterance of B's name prior to or in the absence of the appearance of B itself.

It will be apparent that the Principle of Symmetrical Responding predicts that a child who names the stimuli involved will spontaneously generalize from *With A as sample, pick B* to *With B as sample, pick A* when the symmetry test (Figure 3) is presented. Not quite so obvious is that the Principle of Anticipatory Naming predicts that a child who names the stimuli and in particular names the stimulus B in the presence of A as sample will spontaneously generalize from *With A as sample, pick B* and *With B as sample, pick C* to *With A as sample and in the absence of B as comparison, pick C* when the transitivity test (Figure 5) is presented. This appears when we consider that having learned
  A, pick B, (where A evokes "A" and B evokes "B"),
and
  B, pick C, (where B evokes "B" and C evokes "C"),
the Principle of Anticipatory Naming predicts
  A evokes "A" evokes "B"
whence the standard principles of conditional discrimination learning predict
  "B" evokes pick C (transitivity).
By combining the two principles we can predict that a child who names the stimuli will

also spontaneously generalize from *With A as sample, pick B* to *With A as both sample and comparison and no B, pick A* when the reflexivity test (Figure 2) is presented. Thus, having been trained

A, pick B, (where A evokes "A", and B evokes "B")

the Principle of Anticipatory Naming predicts the spontaneous emergence of

B evokes "B" evokes pick A (symmetry)

while the Principle of Symmetrical Responding predicts the spontaneous continuation

A evokes "A" evokes "B"

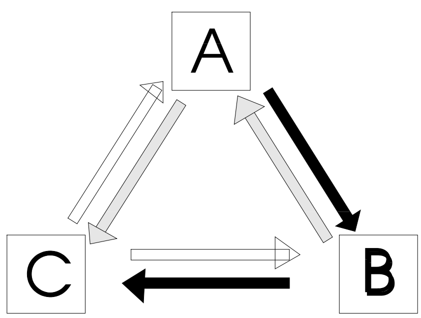
whence the Principle of Anticipatory Naming predicts

"B" evokes pick A (reflexivity).

Predicting Spontaneous Generalization on other Tests of Equivalence

What this shows us is that, taken together, the Principles of Symmetrical Responding and Anticipatory Naming allow us to predict that a child who names the stimuli involved will spontaneously generalize on all three of the tests, reflexivity, symmetry and transitivity, which define an equivalence class. There are, however, two other relations to which the subject's response should (on the assumption that the subject is treating the stimuli as equivalent) and does in fact generalize as a consequence of training A, pick B and B, pick C. These two further relations (C, pick B and C, pick A) appear as the two white arrows on the Classical Equivalence Triangle (Figure 6) in which the two trained relations (A, pick B and B, pick C) are shown as black arrows and the relations tested in the symmetry and transitivity tests (B, pick A and A, pick C) are shown as grey arrows. Since these further relations have sometimes been used by experimenters as tests of equivalence, we need to show that a child who names the stimuli in question would be expected to generalize spontaneously on these tests also. Of these two further relations the response C, pick B after training to respond B, pick C (Figure 4) is predicted by the Principle of Symmetrical Responding in exactly the same way that it predicts the result of the first symmetry test (Figure 3 - B, pick A) after the first training session (Figure 1 - A, pick B).

Figure 6. The Classical Equivalence Triangle



The other case (C, pick A after training A, pick B and B, pick C) is another case like the reflexivity test where spontaneous generalization is predicted from both principles in combination. Thus, having learned

A, pick B, (where A evokes "A" and B evokes "B"),

and

B, pick C, (where B evokes "B" and C evokes "C"),

the Principle of Anticipatory Naming, together with the standard principles of conditional discrimination learning, predicts

A evokes "A" evokes "B" evokes pick C (transitivity),

whence by the Principle of Symmetrical Responding

C, pick A.

It would, however, be a mistake to suppose that reflexivity, symmetry and transitivity, together with the other two relations depicted on the Classical Equivalence Triangle (Figure 6) are the only forms of spontaneous generalization to be observed in the matching-to-sample experiment. There are in fact three other relations emerge from what we may call the Combination Equivalence Triangle depicted on Figure 7. This triangle provided the experimental design used in the research (Lowe and Beasty, 1987; Beasty, 1987) in which the role of naming in mediating the formation of a Sidman stimulus equivalence class in young children was first observed and experimentally manipulated. It differs from the Classical Equivalence Triangle (Figure 6) in that what is trained in the second training session (after the initial A, pick B training) is A, pick C rather than B, pick C. This form of the experiment means that after training A, pick B and testing for reflexivity (A, pick A) and symmetry (B, pick A), no true test of transitivity is possible. As in the case of the Classical Triangle, there is the opportunity, seldom if ever taken in practice, for a second symmetry test (C, pick A) whose outcome, needless to say, is directly predicted by the Principle of Symmetrical Responding. There are also two "combination tests,"[4] those where the equivalence hypothesis predicts B, pick C and C, pick B both of which are sampled in the Lowe and Beasty experiments. Positive results on these tests also are predicted by the naming hypothesis. For having trained

A, pick B, (where A evokes "A" and B evokes "B"),

and

A, pick C, (where A evokes "A" and C evokes "C"),

a combination of the Principle of Symmetrical Responding with the Principle of Anticipatory Naming predicts

B evokes "B" evokes "A",

and

C evokes "C" evokes "A"

whence the standard principles of conditional discrimination learning predict

"A" evokes pick C

and

"A" evokes pick B.

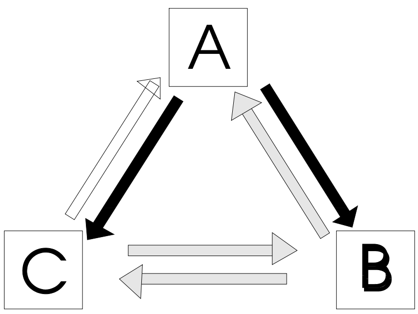


Figure 7. The Combination Equivalence Triangle

Conclusions

If this analysis is correct, the formation of stimulus equivalence classes on the matching-to-sample task by children in the early stages of linguistic development cannot be taken as evidence that at this stage in its development the child already possesses the concept of an equivalence relation. That conclusion would only be justified if spontaneous reflexivity, symmetry and transitivity matching could be shown to depend on a prior ability

(a) to construct some kind of sentence describing the relation between the stimuli involved, and

(b) to draw the appropriate *verbal* inferences from that sentence.

While specific evidence relating to this point is yet to be collected, what evidence there is suggests that equivalence class formation develops in young children long before that level of linguistic sophistication is attained.

Moreover, apart from the unique training in symmetrical responding which provided by simultaneously learning both to select the appropriate name when presented with an instance to which it applies *and* to select the appropriate instance when presented with the name, there is nothing specifically linguistic about this training regime. If this is correct, it would seem to follow that any living organism, a pigeon for example, which been has trained to `name' the stimuli in this sense and has the opportunity to exercise that ability in the test situation ought to show the same pattern of spontaneous generalization on the matching-to-sample task. This is a prediction which is susceptible to experimental test.[5]

One final point. This explanation of the phenomenon assumes that an actual utterance of the name associated with a particular stimulus type is required in order to mediate the spontaneous emergence of equivalence responding. While the evidence supports the view that this mediation of generalization by an explicit naming response is what happens in the case of a child which is in the early stages of acquiring linguistic competence, we may expect that, with repeated experience of the utility of this type of associative generalization, the generalization response will become contingency-shaped and the mediating utterance of the name will tend to drop out. It follows from this

(a) that evidence from the self-reports of older children and adults to the effect that no such no such naming of the stimuli occurs in their case cannot be taken as falsifying the naming hypothesis as applied to the formation of a Sidman stimulus equivalence class in the case of children who are in the early stages of acquiring linguistic competence, and

(b) that the focus of any research designed to test this or any alternative hypothesis concerning the origin of stimulus equivalence responding will need to concentrate on the behavior of children for whom the naming response is in the early stages of acquisition.

REFERENCES

Adams, B. J., Fields, L., and Verhave, T. (1993) Effects of test order on intersubject variability during equivalence class formation. *The Psychological Record, 43,* 133-152.

Beasty, A. (1987). The role of language in the emergence of equivalence relations: A developmental study. Unpublished Ph.D. thesis, University of Wales, Bangor, U.K.

Chomsky, N. (1957) *Syntactic Structures.* 's Gravenhage: Mouton.

Chomsky, N. (1959) Review of B. F. Skinner's *Verbal Behavior. Language*, 35, 26-58.

Crowther, L. M., Dugdale, N. and Lowe, C. F. (1991) How does naming promote stimulus equivalence? The effects of relational and non-relational instructions during name training. Poster presented at the Annual Conference of the Experimental Analysis of Behaviour Group, University College, London.

Dugdale, N. and Lowe, C.F. (1990) Naming and stimulus equivalence. In D. E. Blackman and H. Lejeune (eds.) *Behaviour Analysis in Theory and Practice: Contributions and Controversies*. Hove, East Sussex: Erlbaum, Chapter 5, pp. 115-138.

Fields, L., Adams, B. J., and Verhave, T. (1993) The effects of equivalence class structure on test performances. *The Psychological Record, 43,* 697-712.

Hayes, S. C. (1991) A relational control theory of stimulus equivalence. In L. J. Hayes and P. N. Chase (Eds.), *Dialogues on Verbal Behavior*. Reno, NV: Context Press, pp. 19-40.

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

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

Kant, I. (1781/1787/1929). *Kritik der reinen Vernunft.* First and Second Editions, Riga: Hartknoch. English translation by N.Kemp Smith as *Immanuel Kant's Critique of Pure Reason*, London: Macmillan.

Lowe, C. F. and Beasty, A. (1987) Language and the emergence of equivalence relations. *Bulletin of the British Psychological Society*, 40: A42.

McIntire, K. D., Cleary, J. and Thompson, T. (1987) Conditional relations by monkeys: Reflexivity, symmetry and transitivity. *Journal of the Experimental Analysis of Behavior*, 47: 279-285.

Place, U. T. (1992). Eliminative connectionism and its implications for a return to an empiricist/behaviorist linguistics. *Behavior and Philosophy*, **20**, 21-35

Savage-Rumbaugh, E. S. (1986) *Ape Language: From Conditioned Response to Symbol.* New York: Columbia University Press.

Schusterman, R. J. and Kastak, D. (1993) A Californian sea lion (*Zalophus Californianus*) is capable of forming equivalence relations. *Psychological Record.* 43: 823-839.

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

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

Sidman, M. (1990). Equivalence relations: Where do they come from? In D. E. Blackman and H. Lejeune (Eds.), *Behaviour Analysis in Theory and Practice: Contributions and controversies*. Hillsdale, NJ: Erlbaum, pp. 93-114.

Sidman, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes and L. J. Hayes (Eds.), *Understanding Verbal Relations*. Reno, NV: Context Press, pp. 15-27.

Sidman, M. and Tailby, W. (1982) Conditional discrimination *vs.* matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37: 5-22.

Skinner, B. F. (1938) *The Behavior of Organisms*. New York: Appleton-Century.

Skinner, B. F. (1957) *Verbal Behavior*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1966/1969/1988) An operant analysis of problem solving. In B. Kleinmuntz (Ed.) *Problem Solving: Research, Method and Theory*. New York: Wiley. Reprinted as Chapter 6 of Skinner, B. F. *Contingencies of Reinforcement: A Theoretical Analysis*. New York: Appleton-Century-Crofts. Reprinted with peer comments and a reply in A. C. Catania, and S. Harnad (Eds.) *The Selection of Behavior. The Operant Behaviorism of B. F. Skinner: Comments and Consequences*. Cambridge: Cambridge University Press, pp. 218-236.

Skinner, B. F. (1989) The behavior of the listener. In S. C.H ayes (Ed.) *Rule-Governed Behavior: Cognition, Contingencies and Instructional Control.* New York: Plenum, Chapter 2, pp. 85-96.

NOTES

1. 'Supervised learning' is defined by connectionists as learning which results from an error-correcting feedback supplied by an independent teacher each time an output is emitted. Leaving aside such issues as intermittent reinforcement and the role of aversive (disinforcing) as against attractive (reinforcing) consequences, error messages as against correct messages in the language of connectionism, the only important differences between supervised and

operant learning are

(a) that in most cases it is the environment rather than a human teacher which supplies the error-correcting feedback and
(b) that it is the motivational attitude of the organism to the immediate consequences of the output, rather than the teacher/environment, which determines whether those consequences function as an error message (disinforcer) or correct message (reinforcer).

2. As Kant (1781/1787/1929) points out when he contrasts inspecting the outside of a house with watching a passing ship, whereas in the latter case the sequence is irreversible and in one direction only, cases such as the former (place learning) can and frequently do involve reversible sequences, A→B followed by B→A as one retraces one's steps. Animals are undoubtedly familiar with this kind of symmetry. However, this case differs from that of the symbol in that all the stimuli that make up the sequences are naturally-occurring discriminative stimuli (natural signs) and the feature whereby a particular sequence is selected from a range of possible sequences in both directions is absent.

3. This statement requires qualification in the light of the evidence presented by Crowther, Dugdale and Lowe (1991). They mention the case of a child who made the symmetrical response (B pick A after learning A pick B) after having learned a name for Stimulus A, but not one for Stimulus B. She was able, however, to make a verbal identification of Stimulus B by means of the indexical "That one goes with the triangle." They are also able to point to cases where children used a *common* name for both members of a pair of arbitrarily associated stimuli.

It might be argued in this connection that the concept of one thing's "going with" another represents a primitive grasp on the part of the child of the concept of two things being equivalent. However, although it is clear that the relation of going with something else is symmetrical in that if A goes with B, B goes with A, it is not at all clear either that the relation is reflexive, that A goes with itself, or that it is transitive, that if A goes with B and B goes with C, A must also go with C.

4. So-called on the argument that moving from B as sample to pick C and from C as sample to pick B involve combining an implicit move from B or C as sample to pick A (symmetry) with a move from A as sample to pick C or B (trained).

5. Many claims have been made to have demonstrated spontaneous equivalence responding in animals of a variety of different species from chimpanzees trained in sign language production (Savage-Rumbaugh, 1986), monkeys (McIntire, Cleary and Thompson, 1987) and sea lions (Schusterman and Kastak, 1993). But it is argued (e.g., by Horne and Lowe, forthcoming) that in none of these cases is there convincing evidence that generalization to the test situations was truly spontaneous. In other words, it is either unclear whether care was taken to exclude from the pre-testing training regime training to make the correct response in the presence of the sample stimulus or, as in the McIntire *et al.* experiment, clear that such training was indeed part of the pre-testing regime. However, from the present perspective it is arguable that the training such animals receive is no different that from that which every child receives when learning to name the stimuli involved in the experiment.