Master Thesis
Master in Learning in Complex Systems
MALK 5000

2012

Some Issues in Memory Research within Behavior Analysis

Felix Høgnason

Faculty of Health Science
Oslo and Akershus University College

HØGSKOLEN I OSLO
OG AKERSHUS

Acknowledgement

I would like to thank my wife, Bára Denný Ívarsdóttir, and my two beautiful daughters, Bríet and Þyri Stella, for being supportive and endlessly patient all these months. Now it is time to clear the kitchen table for books and journal articles, and reinstate it for its main purpose.

I would like to thank my supervisor professor Erik Arntzen, for his guidance in data collection and writing this thesis, and Docent Garbriela Sigurdardóttir, for helping me with recruiting the participants at the University of Iceland, and provide suitable locations for the experiment.

Abstract for Both Articles

The purpose of Article 1 is to introduce some of the major theoretical contribution of cognitive psychology concerning short-term or working memory. The role of neuroscientific research of enhancing our understanding of the phenomena will also be discussed. Finally, we will show how conditional discrimination technique in the form of delayed matching-to-sample procedures can be used in research on short-term memory tasks.

Article 2 discusses an experiment were thirty adults were randomly assigned to one of three groups in a single-subject withdrawal design, with three experimental conditions, BAB respectively. The purpose was to examine the probability of responding in accordance with equivalence, as a function of different sample comparison delays in a LS conditional discrimination procedure. The values of the delays were chosen so that the results could be compared to studies by Arntzen (2006), Arntzen et al. (2007), and Eilifsen and Arntzen (2011), and the studies on priming by Posner et al. (1969) and Phillips and Baddeley (1971). The results only partly confirms the results of the aforementioned studies. A second purpose was to examine the effect of exposing participants to different stimulus equivalence procedures, with repeated exposure to one of the procedures. The results show that numbers of trials to criterion was significant lower for all the participants in the repeated procedure, and responses in accordance with stimulus equivalence, was three times higher, indicating a clear repetition, or a carry-over effect.

Article 1

Memory and Remembering: Discussion of Different Approaches to Memory Research

Felix Høgnason

Faculty of Health Science

Oslo and Akershus University College

Abstract

Memory is essential in understanding human behavior according to mainstream psychology. Cognitive theories of memory have been the major contributors to the experimental field of memory, in both psychology and neuroscience. The use of metaphors and hypothetical constructs to account for the relationship between behavior and environment is common in these theories, and the concept of memory is an example of such a construct. Memory, as a capacity, serve as independent behavioral cause, and that kind of causality is circular explanation. Behavior analysis is the study of behavior in its own right. Accepted behavior principles are rooted in experimental analysis, were the experimenter have control of all antecedent variables, and consequences that functional relates with those antecedents. Memory, as a thing or a structure somewhere inside people's brain does not exist. However, people behave in a way that can be described as remembering and recalling, and like other behavior, it can be studied in terms of known behavior principles. Delayed matching-to-sample procedure is considered particularly interesting alternative in the study of short-term memory, and some studies with humans reveal, that responses in accordance with stimulus equivalence enhance, when using delayed matching-to-sample procedures comparing to simultaneous matching-to-sample procedures.

Key words: Cognitive psychology, neuroscience, behavior analysis, matching-to-sample, delayed matching-to-sample, stimulus equivalence

Memory and Remembering: Discussion of Different Approaches to Memory Research

Memory is a topic of central interest in mainstream psychology, and is considered by some to be essential in understanding human behavior (Gleitman, Fridlund, & Reisberg, 2004). According to folk psychology, memory is related to things we know and skills we possess, our ability to orientate, and to tell friends from foe. Memory is the basis for how we define our self, and how we feel.

The cognitive orientation in psychology has been a major contributor to memory research and theory building. Cognitive psychology is based on mentalistic view, which refers to internal causal explanation when explaining behavioral phenomenon. According to this view, there are dimensions inside the organism in some sense - psychic, mental, cognitive, hypothetical, or theoretical, which are qualitative different from the dimension in which behavior takes place. The internal causal phenomenon's are characterized as acts, states, mechanisms, representations, expectancies, memory trace, and so forth, a well-known concepts in our daily speech (Moore, 2008).

According to behavior analysis, behavior is a subject matter in its own right. Behavior is not an index of unobservable structures or constructs inside the organism. Causes of behavior are to be looked for outside the organism and the relation between behavior and environmental variables is functional. The impact of behavior analysis research on memory has been remote, considering the scope of the cognitive contribution. Until recently, the experimental procedures and the scientific concepts of behaviorism has been considered non-applicable by mainstream psychologists, to deal with complex human behavior. However, since the research on stimulus equivalence recommence with the contribution of Murray Sidman and his colleagues (See Sidman, 1994), and with the conceptual contribution of David Palmer concerning cognition and memory (See Palmer, 2003), behavior analysts have acquired conceptual tools that will guide them into the realm of complex human behavior studies.

The purpose of this article is to introduce some of the major theoretical contribution of cognitive psychology concerning short-term or working memory. The role of neuroscientific research of enhancing our understanding of the phenomena will also be discussed. Finally, we will show how conditional discrimination technique in the form of delayed matching-to-sample procedures can be used in research on short-term memory tasks.

Cognitive approach to memory research

There are different opinions on what memory is, what function it has and how to study it. Memory as a subject matter is a major topic in cognitive psychology, and from there most of our common ideas and concepts about memory originates (Baddeley, Eysenck, & Anderson, 2009). The cognitive approach in memory research is usually connected to theories influenced by the development of the computer and associated information processing systems and its usage, which offered new concepts to the field, and terms such as *buffer store*, *feedback*, *encoding* and *retrieval* became prominent part of memory researchers vocabulary (Baddeley, 1998).

One of the most recognized models of memory the past decades, is Atkinson and Shiffrin's *modal model* (Baddeley et al., 2009; Gleitman et al., 2004), where memory is divided into sensory memory, short-term memory and long-term memory. Information from the environment is processed by the sensory memory systems, before it is passed on to a temporary short-term memory and then stored in long-term memory. Storage is a well-known exemplification of memory, where memories are stored like files in an archive or books in a library, and later retrieved. The systems memory span range from a fracture of a second to a lifetime, and the storage capacity from tiny buffers stores to the long-term memory systems (Baddeley, 1998).

In the past decades, a more modern and controversial version of the short-term memory component has taken form; the multi-component model of working memory (Baddeley et al., 2009; Cowan, 2008). Here, the short-term memory has been developed further into a more active and process oriented working memory, divided in to a short-term store and a multi-component model that serves complex cognitive activities such as reasoning, learning and understanding. Even the consciousness lies in the working memory (Baddeley, 1998). One of the most central models in this field is Alan Baddeley's model of working memory (Baddeley, 1998, 2000; Baddeley et al., 2009; Baddeley & Hitch, 1974). Originally, this model consisted of three major components - the central executive, the phonological loop and the visuo-spatial sketchpad. The phonological loop is assumed to hold sequences of acoustic or speech-based items. The visuo-spatial sketchpad does the same with visually and spatially encoded items and arrays. The central executive controls the whole system, by selecting and manipulating materials in the subsystems. At later point, the episodic buffer was added as the fourth component, acting as a storage system, holding fixed amount of information in a multi-dimensional code. The episodic buffer acts as a link between the subsystems of working memory and connects them with input from long-term memory and perception.

Cognitive psychology is a label that is applied to variety of mentalistic orientations (Moore, 2008). Mentalistic orientation explains behavior by appealing to internal causal phenomenon from another dimension, like mental or hypothetical mechanisms, constructs or processes. The use of metaphors and hypothetical constructs to account for the relationship between behavior and environment is common, and observations of relations between behavior and environmental events in laboratory settings, usually entail such hypothetical constructs (Donahoe & Palmer, 2004; Moore, 2008). Memory is an example of a phenomenon explained by theories and models, which themselves are made of hypothetical constructs or hypothetical processes. When human beings learn something, the subject matter is *coded* and *restored* in *memory*. At some later point, the subject matter is *retrieved*, *decoded* and used. Memory, then, is a capacity, and memories serve as independent behavioral causes (Palmer, 1991). This metaphor is a special kind of dualism, which often leads to the fallacy where the terms themselves, or the imaginary systems which they stand for - *memory*, *personality* or *sub-consciousness*, are taken to be the cause of behavior. That kind of a causality is circular explanation or tautology, where hypothetical constructs based on observation of relations between behavior and environment, are used at later point to explain these same relations (Holth, 2001).

In science, laws and principles represent regularities that occur across a wide range of observations. In cognitive psychology, theories and models represent processes underlying particular area or phenomenon. Baddeley (1998) states, that such theories and models are like laws and principles, used to summarize knowledge. In addition to their descriptive functions, they connect results from a number of studies, and the use of hypothetical constructs make way for new and productive hypothesis, which can be tested in experiments, to further increase scientific knowledge (Baddeley, 2000; Baddeley et al., 2009; Donahoe & Palmer, 2004). However, when new knowledge no longer supports older models, the models are either abandoned to make way for new models, or they can be further extended. The hypothetical construct of the short-term memory component of the *modal model* and its development in to Baddeley's (1974) more complex multi-component working memory model, is an example of the latter. The danger of constantly extend an older constructs, is that they become extensive, complicated and even incomprehensible. We could end up with a construct that can explain "all" data, and that contravene the claim of falsifiability of scientific principles and hypothesis; that is, the possibility of providing data that contradicts these same scientific principles and hypothesis and contribute to their refutability (Popper, 2002). What further minimizes the usefulness of using hypothetical constructs in memory studies concerns the practical use of principles drawn from experimental analysis. The emphasis on hypothetical constructs in cognitive psychology results in experiments that primarily tests hypothesis that verify these models. Applied usefulness of many of these studies are questionable and these endeavors tend to actuate the idea that hypothetical construct do indeed underlie behavior. Which means that solutions to applied problems concerning memory lies in restoration or repair of some kind of hypothetical constructs which no one can reach directly (Moore, 2008).

Cognitive psychology is a major contributor to memory research, with theories and concepts influenced by information processing systems. Cognitive psychology tends to explain behavior by appealing to internal hypothetical mechanisms, which are based on observation of the same behavior relations they try to explain. Despite such apparent logical flaws, cognitive oriented theories are still the most prominent theories underlying the neuroscientific approach to memory research.

Neuroscientific approach to memory research

Studies in neuroscience use sophisticated brain imaging technology, like positron emissions scanning (PET), functional magnetic resonance imaging (fMRI), or event-related potentials (ERP), to measure brain activity with varying degrees of temporal and spatial resolution (Faux, 2002; Schlund, Hoehn-Saric, & Cataldo, 2007).

PET scan measures regions of radioactivity in the brain. Human participants are administered radioactive isotopes, and brain regions that are assumed active under cognitive task, become differently radioactive for a short time. Brain activity is usually measured when the participant engages in some behavioral task containing multiple trials. The typical scan takes up to 1 min, and the spatial resolution is between 6 mm and 15 mm, which gives an error margin by several millimeters in any direction (Faux, 2002). Functional MRI uses magnetic fields and the magnetic properties of hemoglobin to produce three-dimensional brain images. It takes up to 6-s for oxygen to concentrate in brain regions, showing the most active parts. An fMRI machine scans about every 2-s with a resolution of about 2 mm. ERP's are brain electrical responses (EEG) which detects tiny electrical currents generated by neurons on the surface of the brain (Gleitman et al., 2004).

In the last few decades, neurological brain studies have mostly been based on cognitive orientation. Usually, cognitive neuroscience bases their interpretation on mentalism, which tends to appeal to a hypothetical inner agent. Researchers in cognitive neuroscience infer about unobserved neural mechanisms from overt behavior, and use the advanced brain technology to give a brain location to these unobserved processes. A typical experimental design in cognitive neuroscience involves the following steps (Faux, 2002): First, a treatment task is identified, involving particular cognitive process. Then, a baseline task, identical to the treatment task, but without the cognitive process, is identified. Separate brain scans, repeated within subjects and tasks, is collected during the baseline and the treatment tasks. Then, an average scan is computed for each individual within each task. The average baseline scan results is subtracted from the average treatment scan results, and brain regions with averages statistically significant from zero, will then account for the cognitive process.

As stated earlier, cognitive processes are based on inference, and cannot be directly observed. It is difficult to take a cognitive operation and divide it in to subcomponents, or cognitive atoms, which in turn makes it hard to design a treatment task that differs from a baseline task by only a single brain operation (Faux, 2002). Variable brain function occurs in the various areas of the brain at all times. As to day, brain-imaging procedures are sensitive only to large regional changes in activation, missing smaller regions of activation. Different PET or fMRI images between treatment task and baseline task, assumed to reveal anatomical locations important to particular cognition, manifest more likely different behaviors (Faux, 2002). Moreover, PET and fMRI technology measures blood flow; they are not direct measures of neural activity. The assumption that momentary regional blood flow reflects the most relevant neural regions of a behavior is just an assumption (Faux, 2002). The PET concepts of activation and deactivation, for example, have no clear relation to underlying neural behavior (Faux, 2002). Brain activation and deactivation in some particular brain regions refer to increase and decrease in localized blood flow and metabolism. Assumed deactivated brain region can still have considerable neural activity and therefore be a neural source for particular behavior (Faux, 2002) .

There is a possibility to study brain functions at the same time the organism is showing specific behavior, and there is basically no contrast between the behavior approach on the one side and the neuroscientific approach on the other side, as long as the neuroscience engages in brain studies where brain functions are measured directly (Skinner, 1976). Dickins et al., (2001) studied brain activation during the formation of equivalence relations. Adult subjects underwent fMRI during matching-to-sample (MTS) tests for baseline relations, symmetry, transitivity, and equivalence, in addition to a test of verbal fluency. The results showed that brain activation was similar in all MTS tasks and in the verbal fluency task. However, the verbal fluency task, but not the MTS tasks activated the Broca's area, which is a brain region of the left front lobe considered to be important in the production of language (Gleitman et al., 2004). And further, the brain activation patterns during the equivalence formation resembled brain activation in semantic processing underlying language, and did not involve regions related to sub-vocal articulations of stimulus names (Dickins et al., 2001).

Cognitive psychology is the major theoretical contributor to the research field in neuropsychology, where one of its major goals is to give a brain location to diverse hypothetical structures. From a behavioristic point of view, neuropsychology could complete the overall understanding of the behavior of organism in many ways. There should be no contrast between the behavior approach and neuroscientific approach, as long as the neuroscience engages in studies where brain functions are measured directly. However, there are still some shortcomings to be overcome in their scientific endeavor, some regarding technical limitation and others concerning their theoretical approach (Skinner, 1976).

Behavioral approach

Behavior analysts are interested in the relation between behavior and environment, where the unit of analysis is functionally defined. The interest of the behavior analysts are the study and explanation of behavior as such, not in imaginary constructs of underlying phenomenon. Behavior analysis uses experiments to study variables in the environment that effect behavior, and these variables must be observable. Analysis of behavior can include specific variables, which are covert or private, and there is no essential difference between overt and covert behavior. However, causes of behavior is always to be looked for outside the organism and the same behavioral principles explains all behavior (Hayes & Brownstein, 1986). In behavior analysis, it is important to adhere coherence in the use of concepts and definitions, and therefore, description of the relations between behavior and environment are based on generally accepted behavioral laws (Donahoe & Palmer, 2004).

In experimental analysis, the experimenter must have control of all antecedent variables and all consequences that functional relates with those antecedents (Donahoe, 2004). When conditions do not permit experimental analysis, the scientist must resort to interpretation. However, interpretation must be constrained by experimental analysis. One of science major goal is to enable interpretation of the larger world on the account of experimental analysis in laboratorial settings. Newtonian principles of classical mechanics, based on experimental analyses conducted with balls and swinging pendulums in the laboratory, are later used to describe the motion of objects in the world outside the laboratory. Interpretation is especially important in accounting for human behavior, where contingencies are complex, reinforcement histories unknown, or experimentation impossible for ethical reasons (Palmer, 1991). Many forms of complex human behavior is a domain of interpretation, mainly because unobservable events cannot be subjected to experimental analysis. Donahoe (2004) states that interpretation may appeal to unobserved events if three requirements are met: "(a) The unobserved operant must be of a kind that has been subjected to prior experimental analysis. (b) The antecedents that prevail when the unobserved operant is invoked must include the critical antecedents identified when it was subjected to experimental analysis. (c) The prevailing conditions must contain antecedents known to be present in the history of the individual when the behavior was reinforced or, minimally, that such antecedents were very likely a part of the history of the individual" (p. 85). Thus, the purpose of behavior analysis is to provide a plausible interpretation of complex human behavior, "interpretations that rest only upon principles that have been established independently of the phenomena to be explained" (Palmer, 2003, p. 7).

Memory, as viewed by cognitive psychology, is not an accepted technical term in behavior analysis (Donahoe & Palmer, 2004; Palmer, 1991). Memory as a storage metaphor, where physiological changes serve as representations of the original stimuli, take over the role of stimuli in controlling behavior (Palmer, 1991), and memory, as a thing or a structure somewhere inside people’s brain does not exist; it is an explanatory fiction (Donahoe & Palmer, 2004; Michael, 2004; Moore, 2008; Palmer, 1991, 2003). However, people behave in a way that can be described as remembering or recalling, and like other behavior, it can be studied in terms of known behavior principles. Like all other behavior, recalling and remembering is subject to the organisms learning history and the environmental contingencies when the behavior occurs. Palmer (1991) accounts for two classes of contingencies when analyzing the behavior that is referred to as memory. In the first class of contingency, a behavior is contingent upon a stimulus at one time, and is still under the control of a stimulus from the same stimulus class at another time. In the second class of contingency, a behavior is under the control of a stimulus, which is no longer present when the appropriate behavior appears and the reinforcement is available. The first contingency is a simple stimulus control and the second contingency is an example of problem solving, or precurrent behavior.

In a simple stimulus control with a pigeon, a key pecking is shaped on a single key in an experimental camber. The key is alternately illuminated or turned off for 5 min. A reinforcer follows pecks in the presence of the light on a VI schedule. Extinction is scheduled when the pigeon pecks on the dark key. At first, the pigeon pecks about the same rate whether the key is illuminated or not. After about 20 min or so, the pigeons key pecking begins to differentiate between conditions, pecking more in the presence of the light, and less when the key is dark. After an hour, the pigeon pecks intensively in the presence of illuminated key, but seldom in the presence of the dark key. After a retention interval of one week, when the pigeon is back in the experimental chamber, it starts to peck with a high rate on the illuminated key, but pecks seldom, if at all, when the key is dark (Palmer, 1991; Pierce & Cheney, 2008).

A child that learns that the verbal stimuli 3x3 produces the spoken response NINE, is an example of simple stimulus control with human. The same response could at later time be elicited through generalization (Delaney & Austin, 1998; Palmer, 1991).

Precurrent behavior has been defined as "any response made by the organism that increases the effectiveness of some subsequent behavior in obtaining a reinforcer" (Arntzen, 2006, p. 136; Skinner, 1968). In the contingencies where precurrent behavior occurs, the behavior that leads to a reinforcer is not available to the organism at that moment. The organism has to behave in a way that the behavior, which leads to reinforcer, has the opportunity to appear. Precurrent behavior as such is not reinforced directly, but enhance the probability of another behavior, which will be reinforced. The reinforcer that followed the second behavior, which comes about because of existing contingencies, maintains the precurrent behavior. Facing a problem, where the correct response is unknown, but probably within ones capability, like remember a familiar person's name in a social setting, is an example of using problem-solving strategies. The situation is probably aversive, and relief only obtainable by generating an appropriate response. The individual could attempt to think of where she met the person in question and what they talked about in that setting. Alternatively, she could go through the alphabet, hoping that a correct letter will elicit the correct response (Donahoe & Palmer, 2004; Palmer, 1991).

Donahoe and Palmer (2004), state the conditions which have to be inherent for problem-solving behavior: "(1) A target response, or set of responses, is in the organism's repertoire and can be evoked by one or more stimulus conditions. (2) Discriminative stimuli are present indicating that the response is scheduled for reinforcement. (3) The current complex of discriminative stimuli is not sufficient to evoke the response directly, or stimuli are present that evoke prepotent competing responses" (p. 271).

Precurrent behavior is seldom visible to others than the organism itself. The behavior is private. Private behavior can also be automatic, which means that the organism is not aware of its own behavior. This could be problematic when using functional analysis to find essential independent variables. In some instances, the only way to come about is to use indirect measures when studying such behavior. Indirect measures could include response- or reaction-time, eye-movement, or neurobiological functions. In some instances one could use such technique as "think-aloud", that is; show to others the behavior (self talk) that would else be covert (Arntzen, Halstadtro, & Halstadtro, 2009; Delaney & Austin, 1998).

Covert behavior like reading is usually learned as overt behavior (Palmer, 1991). Precurrent behavior, like remembering something, is probably also learned as overt behavior. A child could be asked to recollect some incident that happened earlier the same day: "What did you do in the kindergarten today?" Mommy asks. If the child does not know what that means the adult comes to its aid in many ways: By talking about the people the child probably met and the things the child probably did. Precurrent behavior can also occur when a person tries to remember something for a short period of time. Repetition is an example of that kind of precurrent behavior.

In behavior analysis, experiments are used to study observable variables in the relation between the environment and behavior. From these experiments, general behavioral laws are inferred. Sometimes the scientist must resort to interpretation because many conditions do not permit experimental analysis, complex human behavior, like remembering and recalling being some of those domains. Interpretations must be constrained by experimental analysis. They should only be extension of established principles and they cannot discover anything new (Palmer, 2003). Interpretations should not be used when empirical study is possible.

Memory Research in Behavior Analysis

Conditional Discrimination and Matching-to-Sample

An organism discriminates between situations when it shows a certain response in one situation but not in another. The three term contingency, antecedent stimuli (S), response (R) and a consequence (C), is the basic unit of discrimination procedure. Which of many possible discriminations an organism demonstrates is conditioned to other antecedent stimuli. A three-term contingency can be conditioned to a fourth term (S^C), a conditioned stimuli.

Matching-to-sample (MTS) procedures are frequently used to establish conditional discriminations (Green & Saunders, 1998; Michael, 2004; Sidman & Tailby, 1982). MTS procedures have been used in experimental settings across theories and experimental traditions, for example, in neuropsychology, pharmacology, cognitive psychology, behavioral analysis, and in educational settings such as teaching procedure (Green & Saunders, 1998; Sidman, 1994).

In his classical study, Blough (1959) used a delayed-form of matching-to-sample procedure with pigeons. An experimental box was equipped with two translucent plastic response keys, with the third aperture between them, as the sample. A feeding tray, stationed below the response keys, presented grain reinforcement for a few seconds. After magazine training, four pigeons were trained by approximation to peck on the response keys. In the experiment, the sample was first illuminated for 1-s by a white light, either flickering or steady. Then, after *n* seconds, the sample disappeared and the two response keys, previously dark, were illuminated, one by a steady light and the other by a flickering light (randomly assigned to each key). These lights were on until the pigeon pecked either key. If the pigeon pecked the key that matched the sample, it gained access to the feeder for a few seconds. If it pecked the key that did not match the sample, it received no food. In either case, a 5-s inter-trial period intervened where the keys were illuminated by a red light (S^{Δ}) and which there were no opportunity to respond. Then, the next trial began.

We distinguish between identity matching and arbitrary matching. When the sample stimulus and the appropriate comparison stimulus have the same physical characteristics, the procedure is referred to as identity matching-to-sample. When the sample stimulus and the comparison stimulus have no physical similarities, the procedure is referred to as arbitrary matching-to-sample (Sidman & Tailby, 1982).

There are several forms of matching-to-sample procedures. Simultaneous matching-to-sample (SMTS); where the sample and the comparison stimuli are displayed at the same time, and delayed matching-to-sample (DMTS), where the sample disappears when the comparison stimuli appears (0 s delay), or the comparison stimuli appears *n* seconds (*n* s delay) after the sample disappears (Arntzen, 2006).

Delayed matching-to-sample. DMTS procedure is considered to be particularly interesting alternative in the study of short-term or working memory (Arntzen, 2006). The results regarding the matching accuracy of increasing retention intervals with both humans and nonhumans are diverging (Arntzen, 2006; Arntzen et al., 2007; Blough, 1959; Sargisson & White, 2001; Urcuioli & DeMarse, 1997; White, 1985; Wixted, 1989). Retention interval in animal studies are normally between 0 and 20 s (Wixted, 1989). The results from these studies usually show that correct responses to comparisons stimuli decreases when retention interval between the sample disappearance and the display of the comparison, increases (Sargisson & White, 2001). However, Blough (1959) in his classical study, reported about some pigeons who showed sample-specific, stereotypical responses in long delay intervals, maintained accurate matching performance. Pigeons produce some initially irrelevant behavior prior to responding to the comparison stimulus when delays exceeds 0 s, and such behavior can be reinforced accidentally (Keller & Schoenfeld, 1950). If such a mediating behavior enhances subsequent matching performance, the behavior is termed precurrent (Arntzen, 2006; Skinner, 1968).

In their study, Kangas, Berry, and Branch (2011) examined a DMTS performance development over a long period of exposure. Six pigeons were exposed to a DMTS task with variable delays (0, 2, 4, 8 and 16-s) for 300 sessions. Overall, there were 1800 total trials, or 3600 trials per retention interval. The measures used to quantify the development of conditional stimulus control under the procedure (percent correct and a log-*d* measures), showed that high level of accuracy emerge relative quickly under the shorter retention intervals, but increases in accuracy under the longer retention intervals did not emerge until 100-150 sessions were reached. After analyzing of errors, suggestions were made that retention intervals induced biases by shifting control from the sample stimulus to control by key-position. That is, the breakdown of stimulus control induced by increased retention intervals transferred control from the sample stimulus to previously observed control by position (i.e. biases observed during initial training of simultaneous MTS performance).

Kangas et al. (2011) emphasized that an assessment of error report will enhance our understanding of performance development and mechanics.

In the titrating-delay matching-to-sample (TDMTS) procedure, the delay between the sample offset and the comparison onset adjusts as a function of the subject's performance; when the matches are correct the delay increases, but decreases when the matches are incorrect.

In series of three experiments, Kangas, Vaidya, and Branch (2010), explored the effects of several procedural variables on performances in TDMTS procedures. Results from Experiment 1 showed higher daily-titrated delay values (indicating improved remembering) when response requirements on the sample-key was increased. Results from Experiment 2 indicated that the subject's performances adjusted the delay values toward the known baseline delay-value levels, regardless of the initial delay-value. If the goal of the researcher was to reveal the full range of the organism's capability under programmed conditions, it would be better to start each session where the previous left off. In Experiment 3, Kangas et al. (2010) manipulated the step size by which delay values were adjusted. The results showed that a larger step (2-s, compared to 1-s) increased both session-to-session variability and within-session range of titrated delay values.

In study with humans, Steingrimsdottir and Arntzen (2011) conducted an experiment in which conditional discrimination procedures were used with a patient diagnosed with Alzheimer's disease. The purpose of the experiment was to compare performance during arbitrary and identity matching-to sample tasks, and to study the effect of different delays (0-, 3-, 6-, and 9-s) on the participant’s responding. The results showed that the Alzheimer patient responded correctly on the simultaneous identity MTS tasks, but did not respond correctly on the arbitrary MTS tasks. Although the participant did respond correctly on the identity MTS tasks with a 0-s delay, and furthermore with the 3-s, 6-s, and 9-s delays, the number of incorrect responses increased with increasing delays. The results also implied the possibility of using DMTS procedures to evaluate the effect of medication given to Alzheimer's patients.

Stimulus Equivalence

In behavior analysis, research on stimulus equivalence is considered important for understanding of complex human behavior, such as concept formation, categorization and remembering (Arntzen, 2006, 2010; Green & Saunders, 1998; Sidman, 1994). Stimulus equivalence refers to particular pattern in responding not directly taught. This pattern seems to emerge in certain contexts for sets of learned relations between stimuli (Sidman, 1994). Stimulus equivalence classes are usually established using MTS procedures, were conditional discriminations are arranged among arbitrarily assigned sets of stimuli (Arntzen, 2006; Wirth & Chase, 2002). In this paper, the focus is on experimental analysis of stimulus equivalence, which is conceptually consistent with the analysis of Murray Sidman and his co-workers (Sidman, 1992, 1994, 2000; Sidman & Tailby, 1982).

When a human child of certain age, learns to relate a picture of a cookie to the spoken word "cookie", and to relate the printed word COOKIE to the spoken word "cookie", one can assume that the child can relate the picture to the printed word on the first attempt. One can also assume that, when a mother says: "Find a cookie!" The child will select one stimulus from among many. A real cookie, a picture of a cookie, or the printed word COOKIE would serve equally well. The child would also learn, that in a different context, when someone says "Eat the cookie!" that these different modalities of the cookie stimulus, are not literally substitutable (Green & Saunders, 1998).

Stimulus classes and stimulus equivalence. Most human beings do not have to learn directly all the relations between stimuli, that belong to the same stimulus class. Many of these relations emerge without being directly taught, and most people learn with ease to substitute one stimulus for another under different conditions. When two or more stimuli have a common effect on the same response class, they are considered to be part of the same stimulus class. Stimulus classes are assumed to form in several ways. They can be a product of primary stimulus generalization, where two or more stimuli control the same response class, because they have certain physical features in common (Green & Saunders, 1998). A functional class is a class of stimuli that do not have physical characteristics in common, but control the same response class (Sidman, 1994). Stimuli in such a class could have different kind of modality; they could be pictures, smell, sound, text, texture and so on. But they all have the same control on a specific response class (Donahoe & Palmer, 2004). Stimulus class that do not share certain physical properties, and do not necessarily serve the same behavioral function, is referred to as equivalence class (Green & Saunders, 1998; Sidman, 1994). Stimulus equivalence refers to how people relate different stimuli to one another, which has never been directly related before. Generally speaking, stimulus equivalence refers to stimulus substitutability (Green & Saunders, 1998; Sidman, 1994).

Training structures and training protocols. There are three different training structures commonly used in stimulus equivalence studies, that is, one-to-many (OTM) or sample-as-node, many-to-one (MTO) or comparison-as-node, and linear series (LS). There is agreement that linear series is the least effective training structure in producing stimulus equivalence, but there is not a agreement in the literature on whether the MTO or OTM training structure leads to better performance during stimulus equivalence tests (Arntzen, Grondahl, & Eilifsen, 2010; Arntzen & Hansen, 2011; Arntzen & Holth, 2000).

A distinction is made between a training structure and a training protocol. There are three different protocols: Simple-to-complex, complex-to-simple, and simultaneous protocol (Imam, 2006). In the simple-to-complex protocol, each baseline relation is trained first, followed by a symmetry test and a transitivity test. In the complex-to-simple protocol, the baseline relations are trained first, followed by an equivalence test and a mixed test for all of the emergent relations. In the simultaneous protocol, all baseline relations are trained first, before testing for any of the emergent relations; symmetry, transitivity, and equivalence.

Fundamentals of stimulus equivalence methodology. In stimulus equivalence research, stimulus classes are usually referred to with numbers, like 1, 2, 3, and so forth, and the class members with letters, like A, B, C, and so forth. This means that the stimuli A1, B1, and C1 are members of the same stimulus class, A2, B2, C2 are members of another, and A3, B3, and C3 are members of the third one, and so on. A single incident with a sample and comparisons stimuli is called a trial. To minimize the probability of the subjects being familiar with the stimuli or the classification of the stimuli, abstract stimuli, randomly assigned to different classes, are often used with adult subjects. To establish an equivalence class, it is necessary to arrange for a minimum of two classes of three stimuli in each class. To avoid selection by exclusion instead of selection by choice, it is better to have three stimuli classes during conditional training (Green & Saunders, 1998; Sidman, 1987).

In a typical MTS trial with humans, a sample stimulus is presented. When the subject response to it, two or more comparison stimuli, an S^D and S^{Δ}'s, are presented in different locations. Which of the comparison stimuli are the defined S^D and the S^{Δ}'s depends on the sample stimulus. In the presence of the sample stimulus A1, the subject chooses the comparison stimulus B1, and not B2 or B3. Moreover, in the presence of the sample stimulus A2, the subject chooses the comparison stimulus B2, and not B1 or B3, and so forth. In the presence of the sample stimulus B1, the subject chooses the comparison stimulus C1, and not C2 or C3. And, in the presence of the sample stimulus B2, the subject chooses the comparison stimulus C2, and not C1 or C3, and so forth. Here, the former comparison stimuli, B1, B2, B3, have become samples for the new set of comparison stimuli, C1, C2 and C3 respectively. The probability for the subject to choose the right comparison stimuli, when a particular sample is displayed will enhance with repeated exposure. In the training phase, a response is usually followed by a consequence, (e.g., in a textual form), scheduled by the experimenter. If the response is in accordance with the experimenter's definition, words like *right*, *excellent*, *good*, and so on, will be presented (e.g., on a computer screen). If the response is wrong, the word *wrong* appears. Then, the next trial begins after a short intertribal interval (ITI). To consistently meet the requirements of these contingencies, the participant must discriminate among sample stimuli successively presented across trials, and among comparison stimuli simultaneously presented within trials (Green & Saunders, 1998). As the training phases elapses, the programmed consequences are gradually decreased, eventuated in no programmed consequences in the last training phase. When the training phase are over and a conditional discrimination have been established, the subjects are tested for if they respond in accordance with the three property that define stimulus equivalence. No consequences are presented in the test. Stimuli are members of an equivalence class when their interrelations in a matching-to-sample task have the properties of reflexivity, symmetry, and transitivity. Transitivity and symmetry can be tested simultaneously in an LS training structure. When a sample A is related to comparison B (AB) in baseline training, and sample B is related to comparison C (BC), a testing that reveals if the sample C is related to comparison A (CA) without explicit training, would evaluate both symmetry and transitivity of AB and BC (Green & Saunders, 1998). Such a test has been called global equivalence test, or just equivalence test (Arntzen, 2006; Sidman & Tailby, 1982).

Reflexivity, symmetry and transitivity. A conditional relation between stimuli (e.g., if A1 than B1) is directly observable by reference to the subjects interactions with the conditional discrimination procedure (Sidman & Tailby, 1982). Equivalence relations between stimuli, on the other hand, must be tested for, and the subject must show the three properties of the equivalence relations: reflexivity, symmetry and transitivity. The proof of reflexivity is generalized identity matching. To determine that the conditional relation is reflexive, one must show that each stimulus bears the relation to itself: If A1, then A1, must hold true, and if B1, then B1, must hold true. Reflexivity is therefore tested by an identity matching procedure that requires the subject to match stimulus A1 to itself, B1 to itself, and so forth. To say that identity is the basis for the performance, and to rule out the possibility that the subject's performance is controlled by one feature of the sample and another feature of the correct comparison, the subject has to match each new stimulus to itself without differential reinforcement or other current instructions. To demonstrate that the relation is symmetric, one must show that both A1B1 and B1A1 hold true. When matching sample A1 to comparison B1, matching sample B1 to comparison A1 is required without further training. A third stimulus, C1, is required to determine whether the relation is transitive. Subject who has learned two conditional relations A1B1, and B1C1, with the comparison in the first serving as the sample in the second, the proof of transitivity is the emergence of a third conditional relation, A1C1 in which the subject matches the sample from the first relation to the comparison from the second.

DMTS and stimulus equivalence. Some studies with humans, on the relationship between stimulus control and DMTS, reveal that responding in accordance with stimulus equivalence enhances when using DMTS procedures comparing to SMTS procedures, or DMTS procedures without retention time (0 s delay) (Arntzen, 2006; Saunders, Chaney, & Marquis, 2005; Vaidya & Smith, 2006). Saunders, Cheney, and Marquis (2005) found that subjects experiencing a 0 s delay, showed higher yields of equivalence responding than subjects experiencing SMTS. In his study, Arntzen (2006) showed the number of participants who responded in accordance with stimulus equivalence, increased as a function of increased delays, and further, that delays of 9 s did not affect equivalence performance. However, Lian and Arntzen (2011) investigated the effects of 3 s and 6 s delays in DMTS, on responding in accordance with stimulus equivalence with children. There were high yields of derived responding in both groups, but the results did not support the superiority of longer delays. As stated earlier, titrating means that retention interval is variable: When the subject chooses the comparison stimulus, defined by the experimenter as the right one, the retention interval increases – but decreases if the subject chooses the stimulus defined as the wrong one. In their study, Eilifsen and Arntzen (2011) found, that exposing participant with titrating delays between 5000 ms and 8000 ms yielded more positive outcome on stimulus equivalence responding, compared to titrating delays between 0 – 3000 ms. A probable explanation could be that the subject uses some kind of precurrent behavior to complete the reinforcement schedule in such procedures. Precurrent behavior, which appears in the training section of the DMTS, seems to enhance the probability of responding in accordance with stimulus equivalence in the test trials, because the same or similar behavior appears in the test trials. Subjects using DMTS, later reported some kind of precurrent behavior during retention time, while subjects using SMTS did not, or in lesser degree (Vaidya & Smith, 2006). The effect of interrupting potential precurrent behavior was clearly stated in Experiment 4 in Arntzen (2006). Participants in that study, exposed to simultaneous matching and 0-s DMTS, failed to respond in accordance with stimulus equivalence when they were exposed to 3-s DMTS with a restriction to engage in mediating behavior during testing.

Conclusion

Memory is considered essential in understanding human behavior in mainstream psychology, where Atkinson and Shiffrin's *modal model* is one of the most recognized models of memory the past decades. Alan Baddeley's (1974) working memory model, is a further refinement of the short-term memory component of the modal model, where hypothetical constructs, like the central executive, the phonological loop, the visuo-spatial sketchpad, and the episodic buffer, plays a prominent part as internal causal phenomenon to behavior.

Neuropsychology can complete the overall understanding of the behavior of organism in many ways. However, the major theoretical contributor to the neuroscientific research field has been cognitive psychology, where hypothetical structures are given brain location.

Behavior analysis is the study of behavior in its own right. Accepted behavior principles are rooted in experimental analysis, were the experimenter have control of all antecedent variables, and consequences that functional relates with those antecedents. Conditions do not always permit experimental analysis. Complex human behavior is one of those domains where the scientist must resort to interpretation, because unobservable events cannot be subjected to experimental analysis. Interpretation is one of science major goal and it must be constrained by experimental analysis.

Memory, as a thing or a structure somewhere inside people's brain, is not an accepted technical term in behavior analysis. However, people behave in a way that can be described as remembering and recalling, and like other behavior, it can be studied in terms of known behavior principles.

Delayed matching-to-sample procedure is considered particularly interesting alternative in the study of short-term memory. Results from animal studies usually show that correct responses to comparison stimuli decrease when retention interval between the sample disappearance and the display of the comparison, increases. Blough (1959) reported about pigeons, maintaining accurate matching performance , when showing sample-specific, stereotypical responses in long delay. Some studies with humans reveal, that responding in accordance with stimulus equivalence enhance when using DMTS procedures comparing to SMTS procedures, or DMTS procedures without retention time (0 s delay) (Arntzen, 2006; Saunders et al., 2005; Vaidya & Smith, 2006). A probable explanation could be that the subject uses some kind of precurrent behavior to complete the reinforcement schedule in such procedures.

References

Arntzen, E. (2006). Delayed matching to sample: Probability of responding in accord with equivalence as a function of different delays. *The Psychological Record, 56*, 135-167.

Arntzen, E. (2010). Om stimulusekvivalens. Teoretiske betraktninger, oppsummering av en del emperier og noen praktiske implikasjoner. In S. Eikeseth & F. Svartdal (Eds.), *Anvendt atferdsanalyse. Teori og praksis. 2. Utgave.* (2 ed., pp. 100-138). Oslo: Gyldendal Akademisk.

Arntzen, E., Galaen, T., & Halvorsen, L. R. (2007). Different retention intervals in delayed matching-to-sample: Effects of responding in accord with equivalence. *European Journal of Behavior Analysis, 8*(2), 177-191.

Arntzen, E., Grondahl, T., & Eilifsen, C. (2010). The effects of different training structures in the establishment of conditional discriminations and subsequent perfomance on tests for stimulus equivalence. *The Psychological Record, 60*, 437-462.

Arntzen, E., Halstadtro, L.-B., & Halstadtro, M. (2009). The "Silent dog" Method: Analyzing the impact of self-generated rules when teaching different computer chains to boys with autism. *The Analysis of Verbal Behavior, 25*, 51-66.

Arntzen, E., & Hansen, S. (2011). Training structures and the formation of equivalence classes. *European Journal of Behavior Analysis, 12*(2), 483-503.

Arntzen, E., & Holth, P. (2000). Equivalence outcome in single subjects as a function of training structure. *The Psychological Record, 50*, 603-628.

Baddelay, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York, N.Y: Academic Press.

Baddeley, A. (1998). *Human memory. Theory and practice. Revised edition.* Needham Heights, MA.: Allyn and Bacon.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2009). *Memory*. Hove and New York: Psychology Press.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York, N.Y: Academic Press.

Blough, D. S. (1959). Delayed matching in the pigeon. *Journal of the Experimental Analysis of Behavior, 2*(2), 151-160.

Cowan, N. (2008). What are the differences between long-term,short-term, and working memory? *Progress in Brain Research, 169*, 323-338.

Delaney, P. F., & Austin, J. (1998). Memory as behavior: The importance of acquistition and remembering strategies. *The Analysis of Verbal Behavior, 15*, 75-91.

Dickins, D. W., Singh, K. D., Roberts, N., Burns, P., Downes, J. J., Jimmieson, P. (2001). An fmri study of stimulus equivalence. *NeuroReport, 12*(2), 1-7.

Donahoe, J. W. (2004). Interpretation and experimental-anlysis: An underappreciated distinction. *European Journal of Behavior Analysis, 5*(2), 83-89.

Donahoe, J. W., & Palmer, D. C. (2004). *Learning and complex behavior*. Richmond, MA: Ledgetop Publishing.

Eilifsen, C., & Arntzen, E. (2011). Single-subject withdrawal designs in delayed matching-to-sample procedures. *European Journal of Behavior Analysis, 12*(1), 157-172.

Faux, S. F. (2002). Cognitive neuroscience from a behavioral perspective: A critique of chasing ghosts with geiger counters. *The Behavior Analyst, 25*(2), 161-173.

Gleitman, H., Fridlund, A., & Reisberg, D. (2004). *Psychology*. New York: W. W. Norton & Company, Inc. .

Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Peron (Eds.), *Handbook of research methods in human operant behavior*. New York: Plenum Press.

Hayes, S. C., & Brownstein, A. (1986). Mentalism, bahavior-behavior-relations and a behavior-analytic view of the purpose of science. *The Behavior Analyst, 9*, 175-190.

Holth, P. (2001). The persistence of category mistakes in psychology. *Behavior and Philosophy, 29*, 203-219.

Imam, A., A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *Journal of the Experimental Analysis of Behavior, 85*(1), 107-124.

Kangas, B. D., Berry, M. S., & Branch, M. N. (2011). On the development and mechanics of delayed matching-to-sample performance. *Journal of the Experimental Analysis of Behavior, 95*(2), 221-236.

Kangas, B. D., Vaidya, M., & Branch, M. N. (2010). Titrating delay matching-to-sample in the pigeon. *Journal of the Experimental Analysis of Behavior, 94*(1), 69-81.

Keller, F. S., & Schoenfeld, W. N. (1950). *Principles of psychology. A systematic text in the science of behavior.* New York: Appleton-Century-Crofts, inc.

Lian, T., & Arntzen, E. (2011). Training conditional discriminations with fixed and titrated delayed matching-to-sample in children. *European Journal of Behavior Analysis, 12*(1), 173-193.

Michael, J. (2004). *Concepts & principles of behavior analysis*. Kalamazoo: Association for Behavior Analysis International.

Moore, J. (2008). *Conceptual foundations of radical behaviorism*. Cornwall-on-Hudson, NY: Sloan Publishing.

Palmer, D. C. (1991). A behavioral interpretation of memory. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (pp. 261-279). Reno, Nevada: Context Press.

Palmer, D. C. (2003). Cognition. In P. Chase & K. Lattal (Eds.), *Behavior theory and philosophy*. New York: Kluwer Academic / Plenum Publishers.

Pierce, D. W., & Cheney, C. D. (2008). *Behavior analysis and learning* (4 ed.). London: Psychology Press.

Popper, K. (2002). *The logic of scientific discovery*. London: Routledge Classics.

Sargisson, R. J., & White, K. G. (2001). Generalization of delayed matching to sample following training at different delays. *Journal of the Experimental Analysis of Behavior, 75*(1), 1-14.

Saunders, R. R., Chaney, L., & Marquis, J. G. (2005). Equivalence class establishment with two-, three-, and four-choice matching to sample by senior citizens. *The Psychological Record, 55*, 539-559.

Schlund, M., Hoehn-Saric, R., & Cataldo, M. F. (2007). New knowledge derived form learned knowledge: Functional-anatomic correlates of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 87*(2), 287-307.

Sidman, M. (1987). Two choices are not enough. *Behavior Analysis, 22*(1), 11-18.

Sidman, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Heyes (Eds.), *Understanding verbal relations* (pp. 15-27). Reno, Nevada: Context Press.

Sidman, M. (1994). *Equivalence relations and behavior: A research story.* Boston: Authors Cooperative, Inc., Publishers.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74*(1), 127-146.

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*(1), 5-22.

Skinner, B. F. (1968). *The technology of teaching*. Englewood Cliffs, New Jersy: Prentice-Hall, inc.

Skinner, B. F. (1976). *About behaviorism*. New York: Vintage Books.

Steingrimsdottir, H. S., & Arntzen, E. (2011). Using conditional discrimination procedures to study remembering in an alzheimers's patient. *Behavioral Interventions, 26,* 179-192.

Urcuioli, P. J., & DeMarse, T. B. (1997). Memory processes in delayed spatial discriminations: Response intentions or response mediation? *Journal of the Experimental Analysis of Behavior, 67*(3), 323-336.

Vaidya, M., & Smith, K. N. (2006). Brief report: Dellayed matching-to-sample training facilitates derived relational responding. *Experimental Analysis of Human Behavior Bulletin, 24*, 9-16.

White, G. K. (1985). Characteristics of forgetting functions in delayed matching to sample. *Journal of the Experimental Analysis of Behavior*, *44*(1), 15-34.

Wirth, O., & Chase, P. N. (2002). Stability of functional equivalence and wtimulus equivalence: Effects of baseline reversals. *Journal of the Experimental Analysis of Behavior, 77*(1), 29-47.

Wixted, J. T. (1989). Nonhuman short-term memory: A quantitative reanalysis of selected findings. *Journal of the Experimental Analysis of Behavior, 52*(3), 409-426.

Article 2

Responding in Accordance with Stimulus Equivalence: The Effect of Different Delays in a Delayed Matching-to-Sample procedure with a Linear Series Training Structure

Felix Høgnason

Faculty of Health Science

Oslo and Akershus University College

Abstract

Thirty adults were randomly assigned to one of three groups in a single-subject withdrawal design, with three experimental conditions, BAB respectively. The purpose was to examine the probability of responding in accordance with equivalence, as a function of different sample comparison delays in a LS conditional discrimination procedure. Participants were exposed to a delayed matching-to-sample procedure (DMTS) with either 1000 ms, 3000 ms, or 5000 ms delay in the B conditions, a simultaneous matching-to-sample (SMTS) in the A condition, and, a repetition to the B condition. The values of the delays were chosen so that the results could be compared to studies by Arntzen (2006), Arntzen et al. (2007), and Eilifsen and Arntzen (2011), and the studies on priming by Posner et al. (1969) and Phillips and Baddelay (1971). These studies showed a reduction of the effect of the sample on the comparison response when the comparison stimulus was introduced about two – three seconds after the removal of the sample stimulus. The result only partly confirms the results of the aforementioned studies. A second purpose was to examine the effect of exposing participants to different stimulus equivalence procedures, with repeated exposure to one of the procedures. The results show that numbers of trials to criterion was significant lower for all the participants in the repeated procedure, and responses in accordance with stimulus equivalence, was three times higher, indicating a clear repetition, or a carry-over effect.

Key words: conditional discrimination, delayed matching to sample, matching to sample, stimulus equivalence

Responding in Accordance with Stimulus Equivalence: The Effect of Different Delays in a Delayed Matching-to-Sample Procedure with a Linear Series Training Structure

In behavior analysis, research on stimulus equivalence is considered important for understanding of complex human behavior. Stimulus equivalence refers to particular pattern in responding not directly taught. This pattern seems to emerge in certain contexts for sets of learned relations between stimuli (Sidman, 1994). A stimulus equivalence experimental procedure typically involves teaching a participant several conditional discriminations that usually are not in that person's behavioral repertoire, and then test for stimulus equivalence among the stimuli that were involved in these newly established stimulus classes (Arntzen, 2010; Green & Saunders, 1998; Sidman, 1994). Stimulus equivalence classes are usually established by using matching-to-sample (MTS) procedures, in which conditional discrimination are arranged between arbitrarily assigned sets of stimuli (Sidman & Tailby, 1982; Wirth & Chase, 2002). Stimulus classes are usually referred to with numbers (i.e. 1, 2, 3, etc.), and the class members with letters (i.e. A, B, C, etc.). This means that the stimuli A1, B1, and C1 are members of the same stimulus class, A2, B2, and C2 are members of another, and A3, B3, and C3 are members of the third. The minimal arrangement of conditional discrimination training necessary to test for stimulus equivalence is two classes with three members each. However, to avoid selection by exclusion instead of selection by choice, it is better to have three stimuli classes during conditional training (Green & Saunders, 1998; Sidman, 1987). Of three comparisons stimuli presented simultaneously, B1, B2, and B3, the selection of B1 is reinforced in the presence of A1, the selection of B2 is reinforced in the presence of A2, and the selection of B3 is reinforced in the presence of A3. Later, when B1, B2, and B3 are presented as samples, C1, C2, or C3 respectively, serve as correct comparison stimulus. Stimuli are members of an equivalence class when their interrelations in a matching-to-sample task have the properties of reflexivity (e.g., if A1B1, A2B2, and A3B3, then A1A1, A2A2, A3A3 etc.), symmetry (e.g., if A1B1, A2B2, and A3B3, then B1A1, B2A2, and B3A3 etc.), and transitivity (e.g., if A1B1 and B1C1, then A1C1, etc.). Transitivity and symmetry can be tested simultaneously, and such a test has been called global equivalence test, or just equivalence test (Arntzen, 2006; Sidman & Tailby, 1982).

Different structural relations of an equivalence class can be described by several parameters, for example, number of stimuli per class and training structures (Fields & Verhave, 1987). Training structures involve differences in how prerequisite conditional discriminations are sequentially presented for the participants, and how stimuli in the stimuli classes are related in the stimulus equivalence procedures. There are three different training structures commonly used in stimulus equivalence studies; one-to-many (OTM) or sample-as-node, many-to-one (MTO) or comparison-as-node, and linear series (LS). There is agreement that linear series is the least effective training structure in producing stimulus equivalence (Arntzen et al., 2010; Arntzen & Hansen, 2011; Arntzen & Holth, 2000; Fields & Verhave, 1987). One can assume that a training structure that gives a lower yield of stimulus equivalence, could more easily detect variation in performance by reducing the probability of a ceiling effect, which could be more prominent under other circumstances (Arntzen & Holth, 1997).

A simultaneous matching-to-sample (SMTS) procedure is most commonly used in stimulus equivalence research, both when it comes to training of the prerequisite conditional discriminations and when it comes to testing for stimulus equivalence (Arntzen, 2006). Usually, each trial in matching-to-sample procedures starts with a response to the sample stimulus, followed by the presentation of the comparisons stimuli. In SMTS, the sample and the comparisons are present at the same time. In delayed matching-to-sample procedure (DMTS), a response to the sample stimulus is followed, either by the disappearance of the sample and presentation of the comparisons (0 s delay), or *n* second delay before the presentation of the comparisons (Arntzen, 2006). The interval between the sample disappearance and the presentation of the comparison stimulus is usually called the retention interval (Arntzen, 2006).

DMTS procedure has often been used to study remembering in animals (Blough, 1959; Kangas et al., 2011; Kangas et al., 2010; Sargisson & White, 2001; Urcuioli & DeMarse, 1997; White, 1985). The retention in studies involving pigeons range from 0 – 20 seconds, and usually, the matching accuracy declines with increasing delay intervals (Arntzen, 2006; Wixted, 1989). However, in his classical study Blough (1959) discovered that some pigeons maintained accurate matching performance in long delay intervals when they performed some sample-specific, stereotypical responses. Such mediating behavior could be interpreted as a rehearsal during the delay interval. Mediating behavior appears when delays are longer than 0-s, as the probability for the pigeons to engage in initially irrelevant behavior prior to responding to the comparison stimulus at that time is likely (Arntzen, 2006). Such behavior can be accidently reinforced (Keller & Schoenfeld, 1950). Mediating behavior is termed precurrent if it improves subsequent matching performance, and is defined as "any response made by the organism that increases the effectiveness of some subsequent behavior in obtaining a reinforcer" (Arntzen, 2006, p. 136; Skinner, 1968).

Results regarding the matching accuracy of increasing retention intervals has been diverging (Arntzen et al., 2007). Results from some studies with humans indicate, that using delays during training can increase the participant's probability of responding in accordance with stimulus equivalence (Arntzen, 2006; Arntzen et al., 2007; Saunders, Chaney, & Marquis, 2005). And longer delays further enhance responding in accordance with stimulus equivalence than shorter delays (Arntzen, 2006). In Experiment 1 in Arntzen (2006), the number of participants who responded in accordance with stimulus equivalence increased as a function of increased delays for those who started with SMTS procedure in an MTO training structure. In Experiment 2 and 3 in the same study, delays of 9-s did not affect the equivalence performance, using OTM training structure. All of the participants in these experiments responded in accordance with equivalence in all conditions. Saunders, Chaney, and Marquis (2005) implemented two experiments, where senior citizens were trained in 18 sets of conditional discriminations. Training included 2-, 3-, and 4-choice SMTS configurations in LS, MTO and OTM training structures. The experiments aimed to test the assumption that 3- and 4-choice MTS would increase the probability of class establishment, relative to 2-choice MTS, by reducing the possibility of sample/S- control to arise during training. In Experiment 2, six senior citizens participated in a 0-s DMTS procedure, where the goal was otherwise to replicate Experiment 1. The second experiment was conducted to show that the absence of the sample stimulus during responding to the choice stimuli would make acquisition of the conditional discriminations more difficult and possibly have a negative impact on equivalence class establishment. Contrary to their hypothesis, the results showed that trials to criterion for testing were generally fewer in Experiment 2 than in Experiment 1, both for training structure types and for training sets clustered by number of choice stimuli per trial. And the number of training sets that led to criterion performance on tests for equivalence classes were higher in Experiment 2 than in Experiment 1. Vaidya and Smith (2006) used retention intervals of 0 ms, 2000 ms and 8000 ms in both training and testing. Results showed that participants with longer delays were more likely to respond in accordance with symmetry on the tests. The purpose of the study of Arntzen et al. (2007), was to replicate and extend the findings of Arntzen (2006). Twenty participants were exposed to increasing and decreasing retention intervals of 0-s, 6-s, and 12-s delays, and then tested for responding in accordance with stimulus equivalence, using OTM training structure with DMTS procedure. Half of the participants started with 0-s delay with increasing delays, and the other half started with 12-s delay with decreasing delays. A novel set of abstract, arbitrary stimuli was employed for each condition. The results showed that nine out-of 10 participants responded in accordance with equivalence in each conditions, independent of order, and all participants met the criterion for symmetry. Eilifsen and Arntzen (2011), examined the effect of using titrating delays with different starting points during training with an LS structure and simultaneous protocol. In Experiment 1, ten adult participants were exposed to two different conditions, in a variation of a single-subject withdrawal design. Titrating delays between 0 and 3000 ms were used in the A condition, and titrating delays between 5000 ms and 8000 ms in the B condition. Half of the participants were exposed to the conditions in one order, and the other half was exposed to the conditions in the reversed order, ABA and BAB respectively. Results showed that starting the titration of the delay at 5000 ms had positive effects on stimuli equivalence responding for some participants. However, in a series of experiments, Lian and Arntzen (2011) investigated the effects of 3-s and 6-s delays in DMTS, on responding in accordance with stimulus equivalence, using an MTO training structure with children. There were high yields of derived responding in both groups, but the results did not support the superiority of longer delays. Comparing the titrating delayed matching-to-sample procedure with DMTS with fixed delays in this same study revealed that the fixed delay was most effective.

Priming effect is the enduring effect of stimuli on the human senses, after the stimuli have been removed. Priming effects are usually studied by measuring the response latencies to a target stimulus after a priming stimulus has previously been presented (Donahoe & Palmer, 2004). There has been some reports of changes in priming effects when certain delays have been used (Eilifsen & Arntzen, 2011). Some reports shows reduction of the effect on the target response when the target response is introduced about 2000 – 3000 ms after the removal of the priming stimulus. In cognitive oriented literature on priming it has been suggested that this is due to the limits of visualization strategies used in the task (Phillips & Baddelay, 1971; Posner et al., 1969). When delays exceed 2000 – 3000 ms, the participants use other problem solving strategies, which could involve verbal behavior. In a group study by Arntzen, Vaidya, and Eilifsen (2009), results showed that stimulus equivalence yields were lower following a 3000 ms in an DMTS procedure, than performance following both a 100 ms and a 12000 ms delay. Results from this study can said to be consistent with studies on priming, where priming effects have been seen with short retention interval value, like 2000 – 3000 ms (Phillips & Baddelay, 1971; Posner et al., 1969).

It has been argued that DMTS procedure can evoke precurrent behavior and that precurrent behavior can enhance responding in accordance with stimulus equivalence. If that is the case, a distracting task could prevent the possibility of rehearsal of mediating behavior (Arntzen, 2006). Several studies have shown, where participants were engaged in distracting tasks during the retention interval, that none of the participants responded in accordance with stimulus equivalence (Arntzen, 2006). The role of verbal behavior in equivalence formation is one of the major controversies in stimulus equivalence research. Some claim that naming is a prerequisite of emergent relations such as stimulus equivalence (Horne & Lowe, 1996), and that rehearsal of the presented stimuli would yield more positive results on equivalence tests.

Studies have been conducted where the same participant has been exposed to more than one condition for within-subject comparison (Arntzen et al., 2010; Imam, 2006). Only a few studies have been conducted using withdrawal design, where the same participant is exposed to two different stimulus equivalence procedure with repeated exposure to one of the procedures (Eilifsen & Arntzen, 2011). One objection of using single-subject design in stimulus equivalence research is the possibility of carry-over effects from one condition to another. A carry-over effect refers to the influence of the effects of one treatment phase on experimental conditions at a later time (Barlow, Nock, & Hersen, 2009).

Results from studies that have looked at the effect of previous exposure to stimulus equivalence procedures on subsequent stimulus equivalence responding have been inconsistent. In studies employing an LS training structure, the use of simple-to-complex protocol to establish stimulus equivalence classes, increased the possibility on subsequent stimulus equivalence performance following simultaneous protocol training with the same training structure (Buffington, Fields, & Adams, 1997; Fields et al., 1997). A similar study showed, that exposure to a test for derived reasoning that involved transitivity trials, was important presumption for the establishment of stimulus equivalence classes (Fields et al, 2000). This implies a carry-over effect of the previous exposure to stimulus equivalence procedures when certain kind of behavior pattern continues in the unreinforced conditional discrimination task and the test for stimulus equivalence. However, in a study of Holth and Arntzen (2000), an adult participant was exposed to an identical stimulus equivalence procedure involving a simultaneous protocol nine times. Subsequently a new conditional discrimination was trained, using a new set of stimuli but with the same procedure. This participant did not respond in accordance with stimulus equivalence on any of the tests despite the extensive experience the procedures. Pilgrim and Galizio et al. (Pilgrim, Chambers, & Galizio, 1995; Pilgrim & Galizio, 1990) studied the effects of reversing directly trained contingencies on later stimulus equivalence tests. Adult participants were exposed to changed contingencies in the prerequisite conditional discriminations following establishment of stimulus equivalence classes. After a reversal training, which contained the same stimuli as in the previously established stimulus equivalence classes, the participant were tested again for stimulus equivalence. Although a new conditional responding was established and symmetry performance changed according to the reversed contingencies, the responses were consistent with the original stimulus equivalence classes on trials tested for transitivity. These results could imply that some sort of rule-governance, established during the first test, made the participants insensitive to prevailing contingencies during the second test for stimulus equivalence (Eilifsen & Arntzen, 2011; Pilgrim et al., 1995). If this is the case we are probably looking at a carry-over effect which may inhibit stimulus class formation with some participants and not to a lack of extended exposure to stimulus equivalence training and test procedures (Eilifsen & Arntzen, 2011).

An important findings in Experiment 1 in the Eilifsen and Arntzen (2011) study, was that the performance of majority of the participants was similar in all three tests for stimulus equivalence. Experiment 2 in that study was conducted to look directly at the effects of repeated exposure to the same stimulus equivalence procedure. Two participants took part in an identical test and training procedure, where they were exposed to the same titrating delay procedure several times. One participant was exposed to 0 - 3000 ms titrating delay procedure three times, starting with a 0 ms delay in the first training trial, and the other one was exposed to 5000 - 8000 ms titrating delay procedure, starting with a 5000 ms delay in the first training trail. Three different sets of stimuli were used in this experiment, one unique set for each procedure. One of the participants did not respond in accordance with equivalence on any of the tests, but the other one responded in accordance with equivalence on all of the three tests. The results for Experiment 2 appeared to confirm, that previous exposure to the same stimulus equivalence procedure have limited influence on performance in tests for stimulus equivalence in single-subject experimental procedures.

Arntzen (2006) states that higher reaction times during initial testing, may indicate precurrent problem-solving behavior, prior to the selection of a comparison stimulus. Studies have showed differences in reaction time between trials involving directly trained relations, where tests for directly trained relations show higher reaction times for the first test trials, compared to the last training trials and the last test trials (Arntzen, 2006; Arntzen et al., 2010, Eilifsen & Arntzen, 2009; Lian & Arntzen, 2011). These studies have also reported an increase in reaction times on the emergent relations (symmetry and global equivalence) compared to baseline trials and decrease of the reaction times on all types of trials during the course of the test (Arntzen, 2006; Arntzen et al., 2007; Arntzen et al., 2010, Eilifsen & Arntzen, 2009).

The purpose of the current study is to examine the probability of responding in accordance with equivalence in adult human participants, as a function of different sample comparison delays in a LS conditional discrimination procedure. The values of the delays were chosen so that the results could be compared to the studies by Arntzen (2006), Arntzen et al. (2007), and Eilifsen and Arntzen (2011), and the studies on priming by Posner et al. (1969) and Phillips and Baddelay (1971). A second purpose is to examine the effect of exposing participants to different stimulus equivalence procedures, with repeated exposure to one of the procedures.

A single-subject withdrawal design was used in this study. All the participants were first exposed to a B condition, with either 1000 ms, 3000 ms or 5000 ms delay, followed by an A condition of simultaneous matching-to-sample procedure, which again was followed by a repetition of the B condition.

Method

Participants and settings

Thirty adult Icelanders participated in the experiment, 20 women and 10 men, in the ages between 19 and 51 years old. The participants were recruited during lectures in an introductory course in behavior analysis at the University of Iceland, and via personal contacts. The participants were informed that they could withdraw from the experiment at any time. When they finished their session, they were paid 2000 Icelandic kroner (approximately $ 18) for their participation.

The participants were randomly assigned to one of three groups, 10 participants in each group. None of the participants had ever participated in stimulus equivalence research before, or had any experience with equivalence concepts or the stimuli used in the experiment. Each participant was debriefed, thanked, and paid, when their involvement was over.

The experiment was conducted in two locations. In one location, the participants were alone in a 3m x 4m room, seated by a desk facing a blank wall, with a window covered with blind on their right side. The other location was an office, 5m x 7m, where the participants sat by a desk, facing a blank wall, and the experimenter sat at the other end of the room, facing the opposite way. No communication took place between the two when the experiment was running.

Apparatus and stimulus material

A Dell portable computer, with 1400 MHz processor and 520 MB RAM was used in the experiment. The computer was equipped with a 15-inch color monitor and a standard mouse-pointing device. A software program, MTS V3.12, designed by Psych Fusion Ltd. in collaboration with Professor Erik Arntzen at Oslo and Akershus University College, was used for stimulus presentation, data collection and to administer programmed consequences to the participants.

Eighteen visual abstract stimuli, divided in to two sets, one for each condition, were used in the experiment. The stimuli measured between 2.2 - 4.5 cm breadth wise and 1.0 – 2.8 cm in height. The stimulus sets are depicted in Figure 1. The numbers above the columns indicate different classes of stimuli, and the letters on the left of each row indicate different members in each class. The stimuli were displayed on the computer screen, black on a white background. The computer screen was divided in to five squares, four in each corner and one in the center. The sample stimulus was always presented in the center of the screen and three comparison stimuli were presented in random corners with one corner blank.

The participants were given physical copies of the stimuli, printed on laminated cards, and asked to categorize the set before the training and after the test in each condition. The stimuli on the printed copies where black on a white background, measuring between 3.5 – 4.8 cm breadth wise, and 2.5 – 2.7 cm in height.

Procedure

A single-subject withdrawal design was used in the experiment. There were three experimental conditions, BAB respectively, were each condition included seven training phases and a test phase. Figure 2 gives an overview of the experimental procedure. Participants assigned to the first group were exposed to 1000 ms delay in condition B, then to simultaneous matching-to-sample (SMTS) in condition A, followed by a repetition of condition B. Participants in the second group were exposed to 3000 ms delay in condition B, then to SMTS in condition A, followed by a repetition of condition B. Participants in the third group were exposed to 5000 ms delay in condition B, then to SMTS in condition A, followed by a repetition of condition B. All testing involved simultaneous matching, where the sample stimulus remained present after presentation of the comparison stimuli.

An LS training structure with a simultaneous protocol was used in the experiment, which means that all AB and BC trials were introduced randomly and mixed (see Figure 2). The six trials presented were: A1B1B2B3, A2B1B2B3, A3B1B2B3, B1C1C2C3, B2C1C2C3, and B3C1C2C3, were the first alphanumeric character in each string represents the sample stimulus on a given trial, and the underlined alphanumeric character indicate the correct comparison stimulus.

Each experiment condition began with a categorization task, were the participants sorted the laminated stimuli into subject-defined categories. Thereafter, the participants sat down in front of the computer. First, a sample stimulus appeared in the middle of the screen. The participant responded to this stimulus by clicking on it using the right button on the computer mouse. This would make the sample disappear in the B condition, and the comparison stimuli appear in an *n* seconds. In the A condition (SMTS) the sample would remain visible at the same time as the comparison stimuli. If the participant responded in accordance with the stimulus classes defined by the experimenter, the English words "correct", "excellent", "super", "great" or "very good" appeared in the middle of the screen. If the participant responded to a comparison stimulus not defined in the same class as the sample, the word "wrong" appeared in the middle of the screen. At the same time as the programmed consequences appeared, a number indicating the sum of correct responses was displayed in the right bottom corner of the computer screen.

Programmed consequences were provided on all trials in phase one, two and three, on 75% of all trials in phase four, on 50% of all trials in phase five, and on 25% of all trials in phase six. No consequences were provided on the training trials in phase seven or in the test trial. Consequences indicating both correct and incorrect choices lasted for 1000 ms. The inter-trial-interval in both training and testing was 500 ms and the pointing device cursor was reset to a fixed position after each trial.

Throughout the experiment the reaction time to both the sample stimulus and the comparison stimuli were recorded, along with which comparison stimuli the participants responded to.

At the end of test for emergent relation, the participants were re-exposed to the categorization task.

Training and testing

Phases one, two, and three consisted of three repetition of each trial type, where AB and BC relations were trained respectively, resulting in nine trials in the first two phases, and 18 trials in the third phase, where the AB and the BC relations were mixed. Each new phase was started when the participant responded in accordance with defined stimulus classes on more than 90% of trials, eight out of nine trials in phases one and two, and 16 out of 18 trials in the remaining training phases (phases three to seven). If the participant did not reach the criteria, the same phase would continue for another training block.

In the test, direct-trained trials were intermixed with randomized presentation of trials testing for symmetry, transitivity and global equivalence. All test trials were repeated three times, resulting in 56 trials.

Stimulus equivalence responding was defined as responding in accordance with every relation tested for, in 90% or more of the trials testing for each relation. For the performance to be considered as an example of stimulus equivalence, the participants had to respond in accordance with each of the defined relations; directly trained relations, symmetry relations and the global equivalence test, in at least 16 out of 18 trials.

In all the conditions, simultaneous presentation, 1000 ms delay, 3000 ms delay and the 5000 ms delay, the AB and the BC training consisted of nine successive correct responses each and the mixing of the AB and BC trials of 90 successive correct responses, as shown in Table 2. The test block consisted of a random mixing of 54 trials of three types of trials. Eighteen trials tested for directly trained relations, 18 trials tested for responses in accordance with symmetry, and 18 trials tested for responding in accordance with global equivalence. The criterion for responding in accordance with stimulus equivalence was defined as 90% of correct responses or more, in each of the relations, or at least 16 out of 18 tested trials.

General information to the participants

Upon the arrival, the participants were told that the experiment was in the field of experimental behavioral analysis. Then they were required to read through and sign a consent form that informed them about their anonymity and that they were free to terminate their participation at any time. They were told that the length of the experiment was dependent upon how rapidly and correctly they responded. They were also informed that they could take a break at any time during their session, though short break in between the conditions were recommended. After this short briefing, the participants were given physical copies of the stimuli, and asked to categorize the stimuli, anyway they liked. After they completed the categorization task, the participants were instructed to take a seat in front of the computer. When the participants set down in front of the computer, the following instructions[1] were displayed on the screen:

> A stimulus will appear in the middle of the screen. Click on this by using the computer mouse. The stimulus will disappear and three others will appear. Choose one of these by using the computer mouse.
> If you choose the stimulus we have defined as correct, words like very good, excellent, and so on will appear on the screen. If you press a wrong stimulus, the word wrong will appear on the screen. At the bottom of the screen, the number of correct responses you have made will be counted.
> During some stages of the experiment, the computer will not tell you if your choices are correct or wrong. However, based on what you have learned, you can get all the tasks correct.
> Please do your best to get everything right. Good Luck!

Dependent measures

The computer recorded the reaction time, from the presentation of the comparisons to a response to one of the comparisons, and number of trials to criterion as dependent measures. The mean median reaction time for all participants was calculated, for the five last training trials, and for the five first and five last trials testing for directly trained relations, symmetry and global equivalence.

Statistical analyses

For use in statistical analysis, an index of equivalence for each participant was calculated by dividing number of correct responses by the total number of trials in each test phase. Equivalence was defined as an index of 0.9 or higher. Data for training trials and data for responding in accordance with equivalence were analyzed by one-way ANOVA with one group factor (condition) and one dependent variable (index of equivalence and number of

[1] The Instructions were in Icelandic.

training trials). Data for reaction time were analyzed by one-way ANOVA and paired-samples t-test.

Results

As the experiment was run as a withdrawal design, the effect on the dependent variable can be looked at in the same subject. Table 1 shows the main results for every participant, in every group, under all conditions. Participants 9009, 9014, 9015, 9016, 9019, 9020, 9028, 9033, 9037 and 9040 were exposed to the BAB order starting with 1000 ms. Participant 9009 completed training in 253 trials in the B-1 condition, 117 trials in the A condition and 108 trial in the B-2 condition. She responded in accordance with the defining properties of stimulus equivalence in all of the three tests. Participant 9014 completed training in 324 trials in the B-1 condition, 225 trials in the A condition and 108 trials in the B-2 condition. He did not respond in accordance with the defining properties of stimulus equivalence in any of the three tests. He did not respond in accordance with symmetry in the first test, but did so in the second and the third tests. He met the criterion for directly trained relations in the first and the second test, but failed to do so in the third test. Participant 9015 completed training in 612 trials in the B-1 condition, 234 trials in the A condition and 144 trials in the B-2 condition. She did not respond in accordance with the defining properties of stimulus equivalence in the first two tests, but did so in the third test. She did respond in accordance with symmetry and met the criterion for directly trained relations in all the three tests. Participant 9016 completed training in 144 trials in the B-1 condition, 126 trials in the A condition and 108 trials in the B-2 condition. He responded in accordance with the defining properties of stimulus equivalence in all of the three tests. Participant 9019 completed training in 207 trials in the B-1 condition, 117 trials in the A condition and 108 trials in the B-2 condition. She did not respond in accordance with the defining properties of stimulus equivalence or symmetry, nor did she meet the criterion for directly trained relation in any of the tests. Participant 9020 completed training in 243 trials in the B-1 condition, 144 trials in the A condition and 108 trials in the B-2 condition. She responded in accordance with the defining properties of stimulus equivalence in all of the three tests. Participant 9028 completed training in 189 trials in the B-1 condition, 135 trials in the A condition and 108 trials in the B-2 condition. She did not respond in accordance with the defining properties of stimulus equivalence in the first two tests, but did so in the third test. She responded in accordance with symmetry in the second and the third tests and met the criterion for directly trained relations in all the three tests. Participant 9033 completed training in 171 trials in the B-1 condition, 126 trials in the A condition and 108 trials in the B-2 condition. He responded in accordance with the defining properties of stimulus equivalence in all of the three tests. Participant 9037 completed training in 270 trials in the B-1 condition, 162 trials in the A condition and 126 trials in the B-2 condition. She did not respond in accordance with the defining properties of stimulus equivalence or symmetry, nor did she meet the criterion for directly trained relation in the first test, but responded in accordance with equivalence in the second and the third test. Participant 9040 completed training in 594 trials in the B-1 condition, 189 trials in the A condition and 117 trials in the B-2 condition. She did not respond in accordance with the defining properties of stimulus equivalence in any of the three tests. She did respond in accordance with symmetry in the B-1 and the B-2 conditions, but not in the A condition. She met the criterion for directly trained relation in all the tests.

Participants 9001, 9003, 9007, 9018, 9022, 9024, 9030, 9031, 9032 and 9039 were exposed to the BAB order starting with 3000 ms. Participant 9001 completed training in 261 trials in the B-1 condition, 108 trials in the A condition and 108 trial in the B-2 condition. He did not respond in accordance with the defining properties of stimulus equivalence in the first test but did so in the second and the third tests. He responded in accordance with symmetry and met the criterion for direct-trained relations in all the tests. Participant 9003 completed training in 198 trials in the B-1 condition, 162 trials in the A condition and 126 trials in the B-2 condition. She did not respond in accordance with the defining properties of stimulus equivalence in any of the three tests. She responded in accordance with symmetry in the third test and met the criterion for direct-trained relations in the first and the third tests. Participant 9007 completed training in 162 trials in the B-1 condition, 126 trials in the A condition and 108 trials in the B-2 condition. She did not respond in accordance with the defining properties of stimulus equivalence in the first test but did so in the second and the third test. She responded in accordance with symmetry and met the criterion for direct-trained relations in all the three tests. Participant 9018 completed training in 144 trials in the B-1 condition, 153 trials in the A condition and 108 trials in the B-2 condition. She did not respond in accordance with neither the defining properties of stimulus equivalence nor the symmetry in the first test but did respond in accordance with stimulus equivalence in the second and the third test. She met the criterion for direct-trained relations in all the tests. Participant 9022 completed training in 225 trials in the B-1 condition, 117 trials in the A condition and 108 trials in the B-2 condition. He did not respond in accordance with the defining properties of stimulus equivalence in the first test but did so in the second and the third test. He responded in accordance with symmetry and met the criterion for direct-trained relations in all the three tests. Participant 9024 completed training in 144 trials in the B-1 condition, 144 trials in the A condition and 108 trials in the B-2 condition. She did not respond in accordance with the defining properties of stimulus equivalence in the first two tests, but did son in the third test. She responded in accordance with symmetry and met the criterion for direct-trained relations in all the three tests. Participant 9030 completed training in 180 trials in the B-1 condition, 135 trials in the A condition and 108 trials in the B-2 condition. She responded in accordance with the defining properties of stimulus equivalence in all of the three tests. Participant 9031 completed training in 216 trials in the B-1 condition, 117 trials in the A condition and 108 trials in the B-2 condition. She responded in accordance with the defining properties of stimulus equivalence in all of the three tests. Participant 9032 completed training in 126 trials in the B-1 condition, 153 trials in the A condition and 108 trials in the B-2 condition. She did not respond in accordance with the defining properties of stimulus equivalence in the first two tests, but responded in accordance with equivalence in the third test. She responded in accordance with symmetry and met the criterion for direct-trained relations in all the three tests. Participant 9039 completed training in 162 trials in the B-1 condition, 126 trials in the A condition and 108 trials in the B-2 condition. She did not respond in accordance with the defining properties of stimulus equivalence nor did she meet the criterion for direct-trained relations in the first test. She responded in accordance with equivalence in the second and the third test and responded in accordance with symmetry in all the three tests.

Participants 9010, 9012, 9013, 9025, 9026, 9027, 9029, 9034, 9036 and 9038 were exposed to the BAB order starting with 5000 ms. Participant 9010 completed training in 738 trials in the B-1 condition, 171 trials in the A condition and 108 trial in the B-2 condition. He did not respond in accordance with the defining properties of stimulus equivalence or symmetry, nor did he meet the criterion for direct-trained relations in any of the three tests. Participant 9012 completed training in 216 trials in the B-1 condition, 135 trials in the A condition and 108 trials in the B-2 condition. She did respond in accordance with the one node test and met the criterion for direct trained relations in the first test, but did not respond in accordance with symmetry, and therefore did not respond to the defining properties of stimulus equivalence in that test. She responded in accordance with equivalence in the second and the third test. Participant 9013 completed training in 297 trials in the B-1 condition, 180 trials in the A condition and 108 trials in the B-2 condition. He did not respond in accordance with the defining properties of stimulus equivalence or symmetry in any of the three tests, but met the criterion for direct-trained relations in all three tests. Participant 9025 completed training in 189 trials in the B-1 condition, 126 trials in the A condition and 108 trials in the B-2 condition. She did not respond in accordance with the defining properties of stimulus equivalence in the first test, but responded in accordance with symmetry and met the criterion for direct-trained relations in that test. She responded in accordance with equivalence in the second and the third test. Participant 9026 completed training in 216 trials in the B-1 condition, 126 trials in the A condition and 108 trials in the B-2 condition. She did not respond in accordance with the defining properties of stimulus equivalence or symmetry, nor did she meet the criterion for direct-trained relations in the first test, but she responded in accordance with equivalence in the second and the third test. Participant 9027 completed training in 216 trials in the B-1 condition, 126 trials in the A condition and 108 trials in the B-2 condition. He responded in accordance with the defining properties of stimulus equivalence in all of the three tests. Participant 9029 completed training in 162 trials in the B-1 condition, 126 trials in the A condition and 108 trials in the B-2 condition. She did not respond in accordance with the defining properties of stimulus equivalence in the first test, but responded in accordance with symmetry and met the criterion for direct-trained relation in that test. She responded in accordance with equivalence in the second and the third test. Participant 9034 completed training in 162 trials in the B-1 condition, 144 trials in the A condition and 108 trials in the B-2 condition. He responded in accordance with the defining properties of stimulus equivalence in all of the three tests. Participant 9036 completed training in 280 trials in the B-1 condition, 189 trials in the A condition and 108 trials in the B-2 condition. She did not respond in accordance with the defining properties of stimulus equivalence in the first and the third tests, but responded in accordance with symmetry and met the criterion for direct-trained relations in those tests. She responded in accordance with equivalence in the second test. Participant 9038 completed training in 189 trials in the B-1 condition, 270 trials in the A condition and 108 trials in the B-2 condition. He did not respond in accordance with the defining properties of stimulus equivalence in the first test, but responded in accordance with symmetry and met the criterion for direct-trained relations in that test. He responded in accordance with equivalence in the second and the third test.

Another way to analyze the results is to list up the participants in three groups according to different delays in the B condition; the 1000 ms group, the 3000 ms group, and the 5000 ms group. Each session contained three conditions, and each participant was trained and tested in one individual session. The mean session time for the 1000 ms group was 1:46:49 hrs, 1:29:41 hrs for the 3000 ms group, and 1:55:54 hrs for the 5000 ms group. Beside the various programmed delays for each group, the length of each session depended on how rapidly and correctly the participants responded. The participants were given the opportunity to take short breaks between each condition while the experimenter prepared a new condition. The duration of the breaks is not included in the measure of the sessions. All participants completed their session in one day.

Table 1 shows the results for all the groups and all the conditions each group was exposed to. In the B-1 condition, four out of 10 participants met the equivalence criterion in the 1000 ms delay group, two out of 10 participants met the equivalence criterion in the 3000 ms delay group, and two out of 10 participants met the equivalence criterion in the 5000 ms delay group. In the B-1 condition, six out of 10 responded in accordance with symmetry in the 1000 ms delay group, eight out of 10 responded in accordance with symmetry in the 3000 ms delay group and six out of 10 responded in accordance with symmetry in the 5000 ms delay group.

In the A condition, five of 10 participants met the equivalence criterion in the 1000 ms delay group, eight out of 10 participants met the equivalence criterion in the 3000 ms delay group, and eight out of 10 participants met the equivalence criterion in the 5000 ms delay group. In the A condition eight out of 10 responded in accordance with symmetry in the 1000 ms delay group, nine out of 10 responded in accordance with symmetry in the 3000 ms delay group and eight out of 10 responded in accordance with symmetry in the 5000 ms delay group.

In the B-2 condition, seven out of 10 participants met the equivalence criterion in the 1000 ms delay group, nine out of 10 participants met the equivalence criterion in the 3000 ms delay group, and seven out of 10 participants met the equivalence criterion in the 5000 ms delay group. In the B-2 condition, nine out of 10 responded in accordance with symmetry in the 1000 ms delay group, 10 out of 10 responded in accordance with symmetry in the 3000 ms delay group and eight out of 10 responded in accordance with symmetry in the 5000 ms delay group.

The results from statistical analysis on the relation between different groups and responding in accordance with equivalence was not significant under any of the conditions, BAB respectively.

The results of the sorting task are depicted in Figure 3. The figure shows each participant in each group, under all conditions. The bold letters in the EQ column under each condition indicate responding in accordance with stimulus equivalence, and the bold alphanumeric string in each row indicate a participant-defined class that correspond to all of the three experimenter-defined classes. As shown in Table 3, none of the participants categorized the printed copies in the sorting tasks in accordance with any of the experimenter-defined stimulus classes prior to the first and the second condition, B-1 and A, respectively. However, 18 out-of 30 participants categorized the printed copies prior to the third condition (B-2). Twenty out-of 30 participants categorized the printed copies after the first test following the B-1 condition, thereof 13 who did not respond in accordance with equivalence. One participant in the 3000 ms group, who did respond in accordance with stimulus equivalence on the prior test for emergent relation in the B-1 condition, did not categorize the printed copies in the subsequent categorization task. Twenty-three out-of 30 participants categorized the printed copies after the first test following the A condition, thereof four who did not respond in accordance with stimulus equivalence. One participant in the 3000 ms group, who responded in accordance with stimulus equivalence on the prior test for emergent relation in the A condition, did not categorize the printed copies in the subsequent categorization task. Twenty-four out-of 30 participants categorized the printed copies after the first test following the B-2 condition, thereof three who did not respond in accordance with equivalence. Two participants, one in the 1000 ms group and one in the 3000 ms group, who responded in accordance with stimulus equivalence on the prior test for emergent relation in the B-2 condition, did not categorize the printed copies in the subsequent categorization task.

As shown in Figure 3, participants in the group started with the 3000 ms delay have usually lower number of training trials in the first two conditions (B-1 and A, respectively) compared to the 1000 ms and the 5000 ms delay. In the B-2, the difference is unnoticeable. Comparing the groups, participants in the group who started with 1000 ms delay required the highest number of trials to criterion under the B-1 condition, followed by the 5000 ms delay group. The 3000 ms delay group required the lowest number of trials to criterion. Under the A condition the group with the 5000 ms delay in the B conditions required the highest number of trials to criterion, followed by the 1000 ms delay group. The 3000 ms delay group required the lowest number of trials to criterion. Under the B-2 condition, the 1000 ms delay group required the highest number of trials to criterion, followed by the 3000 ms delay group. The 5000 ms delay group required the lowest number of trials to criterion. Statistical analysis of the relation between groups and number of training trials were not significant under any of the conditions.

Comparing the B-1 and the B-2 conditions, a paired-samples t-test show a significant difference in number of trials to criterion for all participants in all groups, where number of trials to criterion are considerable lower in the B-2 condition: $t(29) = 5.38$, $p < .001$. Furthermore, when comparing the B-1 and the A conditions, the paired-samples t-test show a significant difference in number of trials to criterion for all participants in all groups, where number of trials to criterion are considerable lower in the A condition: $t(29) = 4.13$, $p < .001$.

Statistical analysis did not show significant differences in reaction time between groups. Mean reaction time data from the groups are shown in Table 2. In both B conditions (B-1 and B-2) for all groups, there is an increase in reaction time to comparison stimuli from the last five training trials to the first five test trials for the DT relations, and a decrease during testing, from the first five test trials to the last five test trials for all relations. The reaction time for the last five DT test trials is nearly always considerable higher than for the last five DT training trials for all groups under all conditions, except for the 5000 ms group under the A conditions. For all conditions, statistical analysis with one-way-ANOVA showed a significant difference in reaction time between the five last training trails, the five first DT test trials and the five last DT test trials, $F(2, 1347) = 31.8$, $p < .001$. The first five test trials for SYM relations are considerable higher compared to the first five test trials for DT relations for all groups in all conditions, with decreasing reaction time during testing. Moreover, the first five test trials for EQ relations are considerable higher compared to the first five test trials for SYM relation, with decreasing reaction time during testing. Statistical analysis over all conditions, showed significant differences in reaction time for the five first test trials between all relations tested for, DT, SYM, and EQ respectively, $F(2, 1347) = 31.11$, $p < .001$. A paired-samples t-test showed a significant difference in reaction time between the first five test trials, and the last five test trials for SYM relations, for all participants in all conditions, $t(449) = 6.64$, $p <.001$, and a significant difference in reaction time between the first five test trials, and the last five test trials for EQ relations, for all participants in all conditions, $t(449) = 8.45$, $p < .001$

Discussion

The main purpose of this study was to examine the probability of responding in accordance with equivalence in adult human participants, as a function of different sample comparison delays. The values of the delays were chosen so that the results could be compared to the studies by Arntzen (2006), Arntzen et al. (2009), and Eilifsen and Arntzen (2011), and studies on priming by Posner et al. (1969), and Phillips and Baddelay (1971). These studies have showed a reduction of the effect of the sample on the comparison response when the comparison stimulus is introduced about 2000 – 3000 ms after the removal of the sample stimulus. As Table 1 shows, under the B-1 condition, four out of 10 participants of the 1000 ms group responded in accordance with stimulus equivalence, two out of 10 participants of the 3000 ms group responded in accordance with stimulus equivalence, and two out of 10 participants of the 5000 ms group responded in accordance with stimulus equivalence. These results only partly confirm the results from the aforementioned studies as the yields of the 5000 ms group was also expected to be higher than the 3000 ms group.

A second purpose of this experiment was to examine the effect, where the same participant is exposed to two different stimulus equivalence procedures, with repeated exposure to one of the procedures. Comparing the B-1 and the B-2 conditions, a paired-samples t-test show a significant difference in number of trials to criterion for all participants, where number of trials to criterion are considerable lower in the B-2 condition: $t(29) = 5.38$, $p < .001$. Number of participants who responded in accordance with stimulus equivalence is also considerable higher in the B-2 condition than in the B-1 condition, 23 compared to eight, respectively. These differences between these two conditions indicate a clear repetition, or a carry-over effect. When comparing the B-1 and the A conditions, the t-test show a significant difference in number of trials to criterion for all participants, where number of trials to criterion are considerable lower in the A condition: $t(29) = 4.13$, $p < .001$. Number of participants who responded in accordance with stimulus equivalence is also considerable higher in the A condition than in the B-1 condition, 21 compared to eight, respectively. The differences between these two conditions indicate a clear carry-over effect.

For some reason, the carry-over, or repetition effect, is most prominent for the 3000 ms and 5000 ms groups. In this study, seven out of 10 participants in the 1000 ms group responded in congruity on all three tests for stimulus equivalence, either by responding in accordance with stimulus equivalence on all tests, as participants 9009, 9016, 9020, and 9033, or by responding in accordance to none of the relations tested for, like participants, 9014, 9019, and 9040. Only participant 9037, who did not respond in accordance with stimulus equivalence on the first test, did so on the second and the third test. Participants 9015 and 9038, who did not respond in accordance with stimulus equivalence in the first and the second test, did so in the third test. For the 3000 ms group, three out of 10 participants responded in congruity on all three tests for stimulus equivalence, either by responding in accordance with stimulus equivalence on all tests, as participants 9030 and 9031, or by responding in accordance with none of the relations tested for, like participant 9003. Participants 9001, 9007, 9018, 9022, 9032, and 9039, who did not respond in accordance with stimulus equivalence on the first test, did so on the second and the third test. Only participant 9024, who did not respond in accordance with stimulus equivalence in the first and the second test, did so in the third test. For the 5000 ms group, four out of 10 participants responded in congruity on all three tests for stimulus equivalence, either by responding in accordance with stimulus equivalence on all tests, as participants 9027 and 9034, or by responding in accordance with none of the relations tested for, like participants 9010 and 9013. Participant 9012, 9025, 9026, 9029, 9038, who did not respond in accordance with stimulus equivalence on the first test, did so on the second and the third test. Participant 9036, who did not respond in accordance with equivalence on the first test, did so in the second test. However, she did not respond in accordance with stimulus equivalence on the third test, and is the only participant showing negative trend of stimulus equivalence performance in this study. One can clearly see a positive effect on stimulus equivalence performance of previously being exposed to conditional discrimination training and tests of stimulus equivalence for the majority of the participants. In all, eight out-of 30 participants responded in accordance with stimulus equivalence on the first and subsequent tests. Only six out of the remaining 22 participants did not respond in accordance with equivalence on any of the tests, and 16 participants, who did not respond in accordance with stimulus equivalence on the first test, did so in the subsequent tests. One of these six participants, 9003, shows a slow increase in her responding in accordance with direct-trained relations, symmetry relations, and equivalence relations, in the tests throughout the different conditions. The other five seems to confirm the results from the study of Holth and Arntzen (2000), and Experiment 2 in (Eilifsen & Arntzen, 2011), were participants were repeatedly exposed to the same stimulus equivalence procedure, with slight change in stimuli set or the delay, with limited influence on performance on the tests for stimulus equivalence. This could underpin the assumption made by Pilgrim et al. (1995), that for some participant, some sort of rule-governance established during the first test made the participants insensitive to prevailing contingencies during the other tests. The results for the remaining 16 participants, which responded in accordance with stimulus equivalence on the tests after the second and/or third conditions, underpin the assumption that repeated exposures to a stimulus equivalence procedure do enhance responding in accordance with stimulus equivalence on subsequent tests. In this study, the presumed rule-governed behavior is most prominent with the participants exposed to the 1000 ms delay. It would be interesting to explore these findings in later experiments, for example by comparing SMTS with various delays in a DMTS procedure.

Mean reaction time data from the groups are shown in Table 2. There is an increase in reaction time to comparison stimuli from the last five training trials to the first five test trials for the DT relations, and decrease during testing, from the first five test trials to the last five test trials for all relations, in both B conditions (B-1 and B-2), for all groups. The reaction time for the last five DT test trials is nearly always considerable higher than for the last five DT training trials for all groups under all conditions, except for the 5000 ms group under the A conditions. The first five test trials for SYM relations are considerable higher compared to the first five test trials for DT relations, for all groups in all conditions, with the characteristic decrease in reaction time during testing. Moreover, the first five test trials for EQ relations are considerable higher compared to the first five test trials for SYM relation, with the characteristic decrease in reaction time during testing. These results are in accordance with the results from Experiment 1 in Arntzen (2006) and studies of Arntzen et al. (2010), Eilifsen and Arntzen (2011), and, Lian and Arntzen (2011). These studies showed higher reaction times during initial testing, indicating precurrent problem-solving behavior prior to the selection of a comparison stimulus. There were differences in reaction time between trials involving directly trained relations, where tests for directly trained relations showed higher reaction times for the first test trials, compared to the last training trials, and the last test trials. These studies also reported an increase in reaction times on the emergent relations (symmetry and equivalence) compared to baseline trials and decrease of the reaction times on all types of trials during the course of the test. Figures 5, a. - .c. depict the reaction time for six participants, two from each group. In the 1000 ms group, Participant 9009 responded in accordance with stimulus equivalence in all conditions, while Participant 9019 did not respond in accordance with stimulus equivalence in any of the conditions. In the 3000 ms group, Participant 9031 responded in accordance with stimulus equivalence in all conditions, while Participant 9003 did not respond in accordance with stimulus equivalence in any of the conditions. In the 5000 ms group, Participant 9027 responded in accordance with stimulus equivalence in all conditions, while Participant 9010 did not respond in accordance with stimulus equivalence in any of the conditions. Visual inspection reveals that the variance in reaction-time data, between those participants who responded in accordance with stimulus equivalence and those who did not, is evident for the A and B-2 conditions, with greater variance for those participants who did not respond in accordance with stimulus equivalence. This could indicate a continued precurrent behavior for these participants, after the first condition, while the behavior of the participants who responded in accordance with stimulus equivalence became relatively automated in the subsequent conditions.

As would be expected when using the LS structure, the differences in behavior as a function of different experimental manipulations is small. When comparing the 1000 ms group with the 5000 ms group in the B-1 condition, the results is neither in accordance with findings in Experiment 1 in Arntzen (2006) or Experiment 1 in Eilifsen and Arntzen (2011), where higher delays in the prerequisite conditional discrimination training enhanced the possibility of responding in accordance with stimulus equivalence. In fact, participants exposed to 1000 ms delay in the B-1 condition show more stimulus equivalence responding than participants exposed to 5000 ms delay, or four out of 10 compared to two out of 10. However, when the groups are exposed to the A condition the results are quite the opposite, where eight out of 10 participants in the 5000 ms group responded in accordance with equivalence, and only five out of 10 participants in the 1000 ms group responded that way.

In the study of Sunders et al. (2005), the results with the DMTS procedure led to fewer trials to criteria for testing than the SMTS procedure. Participants using DMTS also showed enhanced performance for equivalence class establishment. According to Saunders et al. (2005), a possible explanation could be that the DMTS procedure promoted development of precurrent behavior early in the experiment; the removal of the sample stimulus in the DMTS procedure could have generated naming during the retention interval. Although such precurrent behavior could occur during the SMTS procedure as well, there are possibilities that the DMTS could have prompted such behavior earlier in the experiment using the DMTS procedure, resulting in more training sets leading to class establishment. It has been argued that participants who starts with longer delays uses fewer number of trials to establish prerequisite conditional discriminations, than participants who starts with shorter delays (Eilifsen & Arntzen, 2011). However, as Figure 3 shows, the results are not consistent. As expected, the mean number of trials in the 1000 ms group is higher than the mean number for both the 3000 ms group and the 5000 ms group. One would expect further, that the mean number of trials for the 5000 ms group would be lower than the mean number of trials for the 3000 ms group, but that is not the case here.

A sorting task provides a secondary measure of class formation, according to Fields, Arntzen, Nartey, and Eilifsen (2012). Such a test cannot however, be directly compared with the emergent test for equivalence relation in an MTS procedure, because in the sorting task, the participant can compare all the stimuli in all the classes simultaneously, and freely scan back and forth and so on. This is not possible in the MTS procedure probes. However, a sorting task can indicate the formation of all three classes when the emergent relations tests indicate the same outcome. Under these circumstances, the sorting performances should demonstrate the generalization and maintenance of the equivalence classes across different testing formats, because of the differences between those two tests. In contrast, the sorting task can indicate "partial" class formation when the emergent relations test indicate no class formation. There were only small differences across groups among those participants who sorted the physical copies of stimuli in accordance with the experimenter-defined categories, but did not respond in accordance with stimulus equivalence on the prior test for emergent relations. However, the difference between the first condition (B-1) and the subsequent conditions was apparent; or 13 participant in B-1, versus four in the A condition, and three in the B-2 condition. This could indicate, at least for some of the participants that they were "on the brink of" showing responding in accordance with stimulus equivalence after training and testing in the B-1 condition. What does the sorting task indicate when it does not show the formation of the three classes, but the emergent relations probes indicate those classes? As shown in Figure 4, Participant 9030 responded in accordance with stimulus equivalence on the tests for emergent relation in all conditions, but did not categorize the physical stimuli on any of the subsequent categorization tests. Further studies are required on the role of sorting tasks and sorting task outcomes and especially on their effect in experimental design with repeated exposal's. In summary, because of the diverging results between the emergent relation test and the sorting test, the sorting test cannot be said to show a reliable emergent class formation, as indicated in the study of Fields et al. (2012).

Eight out-of 30 participants in this study, responded accordance with stimulus equivalence after tests in B-1 condition, four out of 10 in the 1000 ms group, two out of 10 in the 3000 ms group, and two out of 10 in the 5000 ms group. This is an average of 2.7 out of every 10 participants. These results are in accordance with earlier findings involving the use of LS training structure with a simultaneous protocol (Arntzen et al., 2010; Arntzen & Holth, 1997; Eilifsen & Arntzen, 2011; Fields et al., 1997). One of the reasons for using the LS structure with simultaneous protocol in this study was to avoid a possible ceiling-effect. However, when using such a deterrent training procedure, the possibility is always at hand that the subsequent test does not clearly detect the difference between conditions, where the differences between the delays, after the disappearance of the sample until the appearance of the comparison, are small. A further refinement on such a training procedure could involve some change regarding the reinforcement magnitude. The study of Odun, Shahan, and Nevin (2005) is especially important regarding the effect of reinforcers on response success in

DMTS. The study shows, like many other similar studies, that the frequency of correct responses is inversely related with the length of the retention interval. The study also revealed that when a sign indicated that the probability of a reinforcer was small (10%), the frequency of correct responses was smaller than when the sign indicated the probability of a reinforcer was high (90%). This was true both when the retention interval was short and when it was long, that is, the frequency of right responses decreased proportionally lesser when the retention interval increased, when a sign indicated that the probability of reinforcers was 90% then when the sign indicated the probability of reinforcer was 10%. The results show that a sign, which indicate reinforce probability, affect the memory of pigeons, and their memory is better if the sign indicates a higher probability of reinforcers. The pigeon's behavior during the retention interval appeared also to be under the influence of the sign, which was visible during the retention interval. When the sign indicated that the probability of reinforcer was high, the pigeons showed more frequent behavior than when the sign indicated a lower probability of reinforcers. This happened even though no visible reinforcement contingencies where operating that should reinforce this kind of behavior. It would be interesting to explore the role of reinforcement magnitude, when using an LS training structure with simultaneous protocol in DMTS procedure, to detect small differences between conditions.

It would be interesting to compare this experiment with an experiment using ABA design, where participants' starts with a SMTS in the A condition, and then be exposed to various delays in the B conditions. Results from that experiment could be compared to similar experiments in Arntzen (2006), and Eilifsen and Arntzen (2011), where results showed that using higher fixed delays during the conditional discrimination training, was more effective in generating stimulus equivalence performance compared to lower delays.

Results from some studies indicate that an DMTS training procedures with relatively high fixed delays facilitate stimulus equivalence responding (Arntzen, 2006). If that is the case, such procedures could be applied as an alternative to SMTS in those cases where stimulus equivalence does not easily emerge. It can be expected that some learners, as children, persons with learning disabilities, or elderly people with dementia in an early phase, could have problems mastering DMTS tasks with high delays. A gradual and individualized approach to high delay, by gradually extend the delay values, could be a way to facilitate stimulus equivalence performance with these learners.

In summary, the result of this study only partly confirms the results of studies by Arntzen (2006), Arntzen et al. (2009), and Eilifsen and Arntzen (2011), and studies on priming by Posner et al. (1969), and Phillips and Baddelay (1971). These studies showed a reduction of the effect of the sample on the comparison response when the comparison stimulus was introduced about two – three seconds after the removal of the sample stimulus. However, the effect of exposing the same participant to two different stimulus equivalence procedures, with repeated exposure to one of the procedures, was clear: Numbers of trials to criterion was significant lower for all the participants in the repeated procedure, and responses in accordance with stimulus equivalence, was three times higher in the same procedure. This indicates a clear repetition, or a carry-over effect. The reaction time for the last five training trials to the first five test trials, and from the first five test trials to the last five test trials, followed a familiar pattern of similar results in studies by Arntzen (2006, Arntzen et al. (2010), Eilifsen and Arntzen (2011), and, Lian and Arntzen (2011). The reaction time was considerable higher for the first five test trials, compared to the last training trials, and decreased as the test elapsed.

References

Arntzen, E. (2006). Delayed matching to sample: Probability of responding in accord with equivalence as a function of different delays. *The Psychological Record, 56,* 135-167.

Arntzen, E. (2010). Om stimulusekvivalens. Teoretiske betraktninger, oppsummering av en del emperier og noen praktiske implikasjoner. In S. Eikeseth & F. Svartdal (Eds.), *Anvendt atferdsanalyse. Teori og praksis. 2. Utgave.* (2 ed., pp. 100-138). Oslo: Gyldendal Akademisk.

Arntzen, E., Galaen, T., & Halvorsen, L. R. (2007). Different retention intervals in delayed matching-to-sample: Effects of responding in accord with equivalence. *European Journal of Behavior Analysis, 8*(2), 177-191.

Arntzen, E., Grondahl, T., & Eilifsen, C. (2010). The effects of different training structures in the establishment of conditional discriminations and subsequent perfomance on tests for stimulus equivalence. *The Psychological Record, 60,* 437-462.

Arntzen, E., Halstadtro, L.-B., & Halstadtro, M. (2009). The "Silent dog" Method: Analyzing the impact of self-generated rules when teaching different computer chains to boys with autism. *The Analysis of Verbal Behavior, 25*, 51-66.

Arntzen, E., & Hansen, S. (2011). Training structures and the formation of equivalence classes. *European Journal of Behavior Analysis, 12*(2), 483-503.

Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record, 47*(2), 309-320.

Arntzen, E., & Holth, P. (2000). Equivalence outcome in single subjects as a function of training structure. *The Psychological Record, 50,* 603-628.

Arntzen, E., Vaidya, M., & Eilifsen, C. (2009). Fixed an titrating delayed matching-to-sample. *Unpublished manuscript*.

Baddelay, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York, N.Y: Academic Press.

Baddeley, A. (1998). *Human memory. Theory and practice. Revised edition.* Needham Heights, MA.: Allyn and Bacon.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2009). *Memory*. Hove and New York: Psychology Press.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York, N.Y: Academic Press.

Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying bahavior change.* (3rd ed.). Boston: Pearson Education, Inc.

Blough, D. S. (1959). Delayed matching in the pigeon. *Journal of the Experimental Analysis of Behavior, 2*(2), 151-160.

Buffington, D., M., Fields, L., & Adams, B. J. (1997). Enhancing equivalence class formation by pretraining of other equivalence classes. *The Psychological Record, 47*(1), 69-96.

Cowan, N. (2008). What are the differences between long-term,short-term, and working memory? *Progress in Brain Research, 169*, 323-338.

Delaney, P. F., & Austin, J. (1998). Memory as behavior: The importance of acquistition and remembering strategies. *The Analysis of Verbal Behavior, 15*, 75-91.

Dickins, D. W., Singh, K. D., Roberts, N., Burns, P., Downes, J. J., Jimmieson, P. (2001). An fmri study of stimulus equivalence. *NeuroReport, 12*(2), 1-7.

Donahoe, J. W. (2004). Interpretation and experimental-anlysis: An underappreciated distinction. *European Journal of Behavior Analysis, 5*(2), 83-89.

Donahoe, J. W., & Palmer, D. C. (2004). *Learning and complex behavior*. Richmond, MA: Ledgetop Publishing.

Eilifsen, C., & Arntzen, E. (2011). Single-subject withdrawal designs in delayed matching-to-sample procedures. *European Journal of Behavior Analysis, 12*(1), 157-172.

Faux, S. F. (2002). Cognitive neuroscience from a behavioral perspective: A critique of chasing ghosts with geiger counters. *The Behavior Analyst, 25*(2), 161-173.

Fields, L., Arntzen, E., Nartey, R. K., & Eilifsen, C. (2012). Effects of a meaningful, a discriminative, and a meaningless stimulus on equivalence class formation. *Journal of the Experimental Analysis of Behavior, 97*(2), 163-181.

Fields, L., Reeve, K. F., Rosen, D., Varelas, A., Adams, B. J., Belanich, J. (1997). Using the stimultaneous protocol to study equivalence class formation: The facilitating effects of nodal number and size of previously established equivalence classes. *Journal of the Experimental Analysis of Behavior, 67*(3), 367-389.

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

Gleitman, H., Fridlund, A., & Reisberg, D. (2004). *Psychology*. New York: W. W. Norton & Company, Inc. .

Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Peron (Eds.), *Handbook of research methods in human operant behavior*. New York: Plenum Press.

Hayes, S. C., & Brownstein, A. (1986). Mentalism, bahavior-behavior-relations and a behavior-analytic view of the purpose of science. *The Behavior Analyst, 9*, 175-190.

Holth, P. (2001). The persistence of category mistakes in psychology. *Behavior and Philosophy, 29*, 203-219.

Holth, P., & Arntsen, 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*(1), 181-241.

Imam, A., A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *Journal of the Experimental Analysis of Behavior, 85*(1), 107-124.

Kangas, B. D., Berry, M. S., & Branch, M. N. (2011). On the development and mechanics of delayed matching-to-sample performance. *Journal of the Experimental Analysis of Behavior, 95*(2), 221-236.

Kangas, B. D., Vaidya, M., & Branch, M. N. (2010). Titrating delay matching-to-sample in the pigeon. *Journal of the Experimental Analysis of Behavior*, *94*(1), 69-81.

Keller, F. S., & Schoenfeld, W. N. (1950). *Principles of psychology. A systematic text in the science of behavior.* New York: Appleton-Century-Crofts, inc.

Lian, T., & Arntzen, E. (2011). Training conditional discriminations with fixed and titrated delayed matching-to-sample in children. *European Journal of Behavior Analysis, 12*(1), 173-193.

Michael, J. (2004). *Concepts & principles of behavior analysis*. Kalamazoo: Association for Behavior Analysis International.

Moore, J. (2008). *Conceptual foundations of radical behaviorism*. Cornwall-on-Hudson, NY: Sloan Publishing.

Odum, A. L., Shahan, T. A., & Nevin, J. A. (2005). Resistance to change of forgetting functions and response. *Journal of the Experimental Analysis of Behavior*, *84*(1), 65-75.

Palmer, D. C. (1991). A behavioral interpretation of memory. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (pp. 261-279). Reno, Nevada: Context Press.

Palmer, D. C. (2003). Cognition. In P. Chase & K. Lattal (Eds.), *Behavior theory and philosophy*. New York: Kluwer Academic / Plenum Publishers.

Phillips, W. A., & Baddelay, A. D. (1971). Reaction time and short-term visual memory. *Psychonomic Science*, *22*(2), 73-74.

Pierce, D. W., & Cheney, C. D. (2008). *Behavior analysis and learning* (4 ed.). London: Psychology Press.

Pilgrim, C., Chambers, L., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: Ii. Children. *Journal of the Experimental Analysis of Behavior, 63*(3), 239-254.

Pilgrim, C., & Galizio, M. (1990). Relations between baseline contingencies and equivalence probe performances. *Journal of the Experimental Analysis of Behavior, 54*(3), 213-224.

Popper, K. (2002). *The logic of scientific discovery*. London: Routledge Classics.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph, 79*(1), 1-16.

Sargisson, R. J., & White, K. G. (2001). Generalization of delayed matching to sample following training at different delays. *Journal of the Experimental Analysis of Behavior, 75*(1), 1-14.

Saunders, R. R., Chaney, L., & Marquis, J. G. (2005). Equivalence class establishment with two-, three-, and four-choice matching to sample by senior citizens. *The Psychological Record, 55*, 539-559.

Schlund, M., Hoehn-Saric, R., & Cataldo, M. F. (2007). New knowledge derived form learned knowledge: Functional-anatomic correlates of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 87*(2), 287-307.

Sidman, M. (1987). Two choices are not enough. *Behavior Analysis, 22*(1), 11-18.

Sidman, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Heyes (Eds.), *Understanding verbal relations* (pp. 15-27). Reno, Nevada: Context Press.

Sidman, M. (1994). *Equivalence relations and behavior: A research story.* Boston: Authors Cooperative, Inc., Publishers.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74*(1), 127-146.

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*(1), 5-22.

Skinner, B. F. (1968). *The technology of teaching*. Englewood Cliffs, New Jersy: Prentice-Hall, inc.

Skinner, B. F. (1976). *About behaviorism*. New York: Vintage Books.

Steingrimsdottir, H. S., & Arntzen, E. (2011). Using conditional discrimination procedures to study remembering in an alzheimers's patient. *Behavioral Interventions, 26*, 179-192.

Urcuioli, P. J., & DeMarse, T. B. (1997). Memory processes in delayed spatial discriminations: Response intentions or response mediation? *Journal of the Experimental Analysis of Behavior, 67*(3), 323-336.

Vaidya, M., & Smith, K. N. (2006). Brief report: Dellayed matching-to-sample training facilitates derived relational responding. *Experimental Analysis of Human Behavior Bulletin, 24*, 9-16.

White, G. K. (1985). Characteristics of forgetting functions in delayed matching to sample. *Journal of the Experimental Analysis of Behavior*, *44*(1), 15-34.

Wirth, O., & Chase, P. N. (2002). Stability of functional equivalence and wtimulus equivalence: Effects of baseline reversals. *Journal of the Experimental Analysis of Behavior*, *77*(1), 29-47.

Wixted, J. T. (1989). Nonhuman short-term memory: A quantitative reanalysis of selected findings. *Journal of the Experimental Analysis of Behavior*, *52*(3), 409-426.

Overview of Tables and Figures

Article 1

No tables or figures

Article 2

Page

Table 1.
Results.
The main results for each participant, under each condition; participants number, gender and age; condition, number of training trails to criterion and number of correct responses to DT, SYM and EQ

P#	*Gen*	*Age*	*Cond.*	*Tr. tr.*	*DT*[a]	*SYM*	*EQ*	*Cond.*	*Tr. tr.*	*DT*	*SYM*	*EQ*	*Cond.*	*Tr. tr.*	*DT*	*SYM*	*EQ*
9009	F	22	1000	252	**18**	**18**	**18**	SMTS	117	**18**	**18**	**18**	1000	108	**18**	**18**	**18**
9014	M	30	1000	324	**17**	16	10	SMTS	225	**18**	**17**	6	1000	108	16	**17**	9
9015	F	21	1000	612	**18**	**17**	12	SMTS	234	**17**	**17**	16	1000	144	**17**	**18**	**17**
9016	M	22	1000	144	**18**	**18**	**18**	SMTS	126	**18**	**18**	**18**	1000	108	**18**	**18**	**18**
9019	F	20	1000	207	13	11	0	SMTS	117	16	15	7	1000	108	16	12	1
9020	F	28	1000	243	**18**	**18**	**18**	SMTS	144	**18**	**18**	**18**	1000	108	**18**	**18**	**18**
9028	F	38	1000	189	**18**	16	6	SMTS	135	**18**	**18**	6	1000	108	**18**	**18**	**18**
9033	M	28	1000	171	**18**	**18**	**18**	SMTS	126	**18**	**18**	**18**	1000	108	**18**	**18**	**18**
9037	F	26	1000	270	14	13	2	SMTS	162	**18**	**18**	**18**	1000	126	**18**	**18**	**17**
9040	F	39	1000	594	**17**	**17**	3	SMTS	189	**17**	16	6	1000	117	**18**	**17**	10
9001	M	20	3000	261	**18**	**17**	10	SMTS	108	**18**	**18**	**18**	3000	108	**18**	**17**	**18**
9003	F	25	3000	198	**18**	15	2	SMTS	162	14	14	8	3000	126	**18**	**18**	16
9007	F	21	3000	162	**17**	**17**	13	SMTS	126	**18**	**18**	**18**	3000	108	**18**	**18**	**18**
9018	F	20	3000	144	**18**	15	10	SMTS	153	**18**	**18**	**17**	3000	108	**18**	**18**	**18**
9022	M	22	3000	225	**17**	**17**	10	SMTS	117	**18**	**18**	**18**	3000	108	**18**	**18**	**18**
9024	F	21	3000	144	**17**	**18**	6	SMTS	144	**18**	**18**	15	3000	108	**18**	**17**	**18**
9030	F	21	3000	180	**18**	**18**	**18**	SMTS	135	**18**	**17**	**17**	3000	108	**18**	**18**	**18**
9031	F	23	3000	216	**18**	**18**	**17**	SMTS	117	**18**	**18**	**18**	3000	108	**18**	**18**	**17**
9032	F	27	3000	126	**17**	**18**	15	SMTS	153	**18**	**18**	**18**	3000	108	**18**	**18**	**18**
9039	F	22	3000	162	15	**18**	14	SMTS	126	**18**	**17**	**18**	3000	108	**18**	**18**	**18**
9010	M	20	5000	738	12	10	3	SMTS	171	12	14	9	5000	108	15	13	6
9012	F	19	5000	216	**18**	16	**17**	SMTS	135	**18**	**17**	**17**	5000	108	**18**	**18**	**18**
9013	M	51	5000	297	**18**	16	11	SMTS	180	**17**	13	9	5000	108	**17**	16	12
9025	F	22	5000	189	**18**	**17**	16	SMTS	126	**18**	**18**	**18**	5000	108	**18**	**18**	**18**
9026	F	26	5000	216	14	14	12	SMTS	126	**18**	**18**	**18**	5000	108	**18**	**17**	**17**
9027	M	22	5000	216	**18**	**18**	**18**	SMTS	126	**18**	**18**	**18**	5000	108	**18**	**18**	**18**
9029	F	21	5000	162	**17**	**17**	11	SMTS	126	**18**	**18**	**18**	5000	108	**18**	**18**	**18**
9034	M	29	5000	162	**18**	**17**	**18**	SMTS	144	**18**	**18**	**17**	5000	108	**18**	**18**	**18**
9036	F	47	5000	280	**18**	**17**	7	SMTS	189	**18**	**18**	**18**	5000	108	**18**	**18**	11
9038	M	21	5000	189	**18**	**17**	16	SMTS	270	**18**	**18**	**18**	5000	108	**18**	**18**	**18**

Note: P# = number assigned to each participant; Gen = gender; Age = age in years; Con. = condition; Tr. tr. = number of training trails to criterion; DT = number of correct responses when tested for direct trained relations; SYM = number of correct responses when tested for symmetry relations; EQ = number of correct responses when tested for equivalence relations.
[a] There where 18 test trails for each relation tested for; direct trained relations (DT), symmetry relations (SYM), and equivalence relations (EQ)

Table 2.
The Mean Reaction Time for all Groups under all Conditions

			1000 ms group			*3000 ms group*			*5000 ms group*		
			5 last training	5 first test	5 last test	5 last training	5 first test	5 last test	5 last training	5 first test	5 last test
Conditions	B1	DT	1,58	5,69	3,05	1,31	2,70	2,17	1,63	2,91	2,01
		SY		4,38	2,87		3,77	2,86		4,61	2,21
		EQ		8,44	5,16		5,89	3,96		9,05	4,10
	A	DT	1,74	2,38	2,22	1,56	2,18	1,87	1,86	1,78	1,80
		SY		3,00	2,31		3,77	2,01		2,85	1,80
		EQ		5,63	3,84		4,38	2,50		3,78	2,85
	B2	DT	1,50	2,38	2,23	1,19	1,79	1,76	1,49	2,04	1,56
		SY		2,93	2,54		2,60	1,87		2,29	1,62
		EQ		5,73	4,05		2,92	2,39		3,95	2,55

The table shows the mean reaction time for the five last training trials, the five first test trials and the five last test trials for direct trained relations (DT), and, the five first test trials and the five last test trials for symmetry (SYM) and stimulus equivalence (EQ) relations.

Table 3.
Sorting task outcome and responding in accordance with EQ

			Groups		
			1000 ms	3000 ms	5000 ms
Conditions	B-1	Prior	0	0	0
		EQ	4	2	2
		After	**0** / 4 / 1	**1** / 1 / 6	**0** / 2 / 6
	A	Prior	0	0	0
		EQ	5	7	8
		After	**0** / 5 / 3	**1** / 6 / 1	**0** / 8 / 0
	B-2	Prior	**2** / 5 / 0	**4** / 5 / 1	**0** / 7 / 0
		EQ	7	9	7
		After	**1** / 6 / 1	**1** / 8 / 1	**0** / 7 / 1

The table shows the number of participants who responded in accordance with equivalence (EQ row) in all groups under all conditions, prior to training and after the tests for emergent relations. In the three-number columns: The middle number indicates the number of participants who responded in accordance with stimulus equivalence and sorted the laminated stimuli into participant-defined classes that corresponded to all three experimenter-defined classes. The bold number to the left indicates the number of participants who responded in accordance with stimulus equivalence, but did not sort the laminated stimuli in accordance with all three experimenter-defined classes. The number to the right indicates the number of participants who did not respond in accordance with stimulus equivalence but sorted the laminated stimuli with all three experimenter-defined classes.

Figure 1.
The Stimuli Used in the Experiment

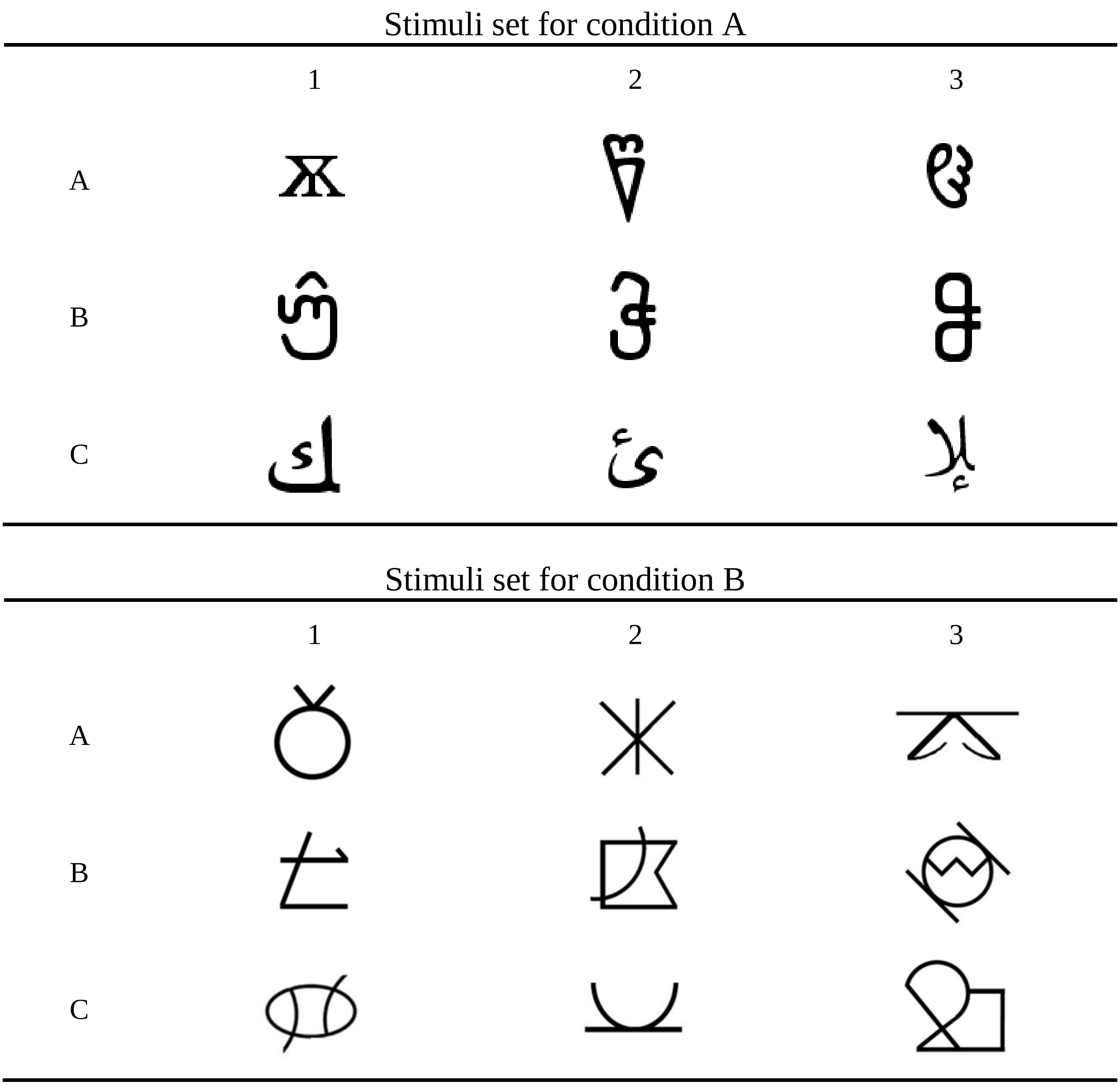
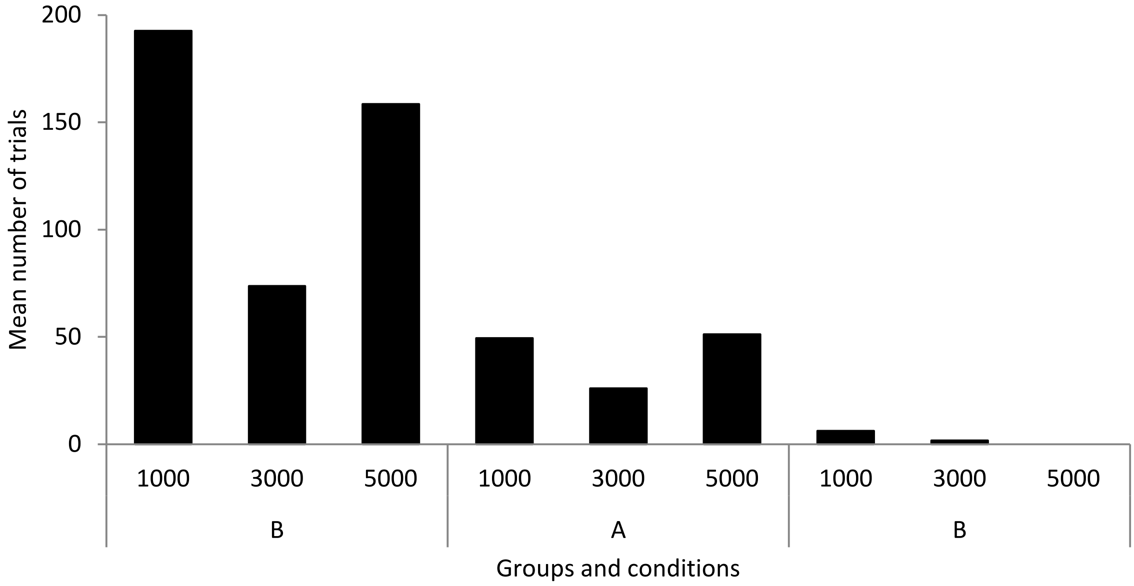


Figure 2:
The Procedure of the Experiment

	Condition description	Phases	Trial types	Feedback	Minimum # of trials
B	Participant exposed to either a **1000 ms** **3000 ms** **5000 ms** delay between the disappearance of the sample and the appearance of the comparisons during training	**Training phase 1: AB training** AB trials presented randomly	A1B1, A2B2, A3B3	100%	9
		Training phase 2: BC training BC trials presented randomly	B1C1, B2C2, B3C3	100%	9
		Training phase 3: Mixed training All training trials presented randomly	A1B1, B1C1, A2B2, B2C2, A3B3, B3C3	100 %	18
		Training phase 4: Feedback fading 1 All training trials presented randomly	A1B1, B1C1, A2B2, B2C2, A3B3, B3C3	75%	18
		Training phase 5: Feedback fading 2 All training trials presented randomly	A1B1, B1C1, A2B2, B2C2, A3B3, B3C3	50%	18
		Training phase 6: Feedback fading 3 All training trials presented randomly	A1B1, B1C1, A2B2, B2C2, A3B3, B3C3	25%	18
		Training phase 7: Feedback fading 4 All training trials presented randomly	A1B1, B1C1, A2B2, B2C2, A3B3, B3C3	None	18
		Test phase All types of test trial were introduced mixed and randomly	A1B1, B1C1, A2B2, B2C2, A3B3, B3C3 (DT) B1A1, C1B1, B2A2, C2B2, B3A3, C3B3 (SY) A1C1, C1A1, A2C2, C2A2, A3C3, C3A3 (1N)	None	54
A	Participants exposed to simultaneous match-to-sample during training		Identical to condition B		
B	Repeated condition		Repeated condition		

Figure 3.
Training trials above the 108-trials minimum as a function of different delays



The figure shows the average number of training trials above the 108-trials minimum as a function of different delays.

Figure 4.

Sorting task outcome and responding in accordance with equivalence.

Condition B-1

P#	Delay ms	EQ	Pre Class Formation					Post Class Formation			
9009	1000	**18/18**	A1B1C2B2A2C3A3C1B3					**A1B1C1**	**A2B2C2**	**A3B3C3**	
9014	1000	10/18	A1B1C3	C1C2B2	B1A3A2			B3A3	B1C1A1	B2A2	C3C2
9015	1000	12/18	A1C1C3B3A2A3C2B2B1					**C2B2A2C3B3A3A1B1C1**			
9016	1000	**18/18**	A1C1B3	A3C2B1	A2C3B2			**A1B1C1**	**A3B3C3**	**A2B2C2**	
9019	1000	0/18	C1C2B3A1A3C3B2B1A2					B3C3	B2C2	A3A1A2	B1C1
9020	1000	**18/18**	B2A1A2	C2B1C1A3C3B3				**B1C1A1**	**B2A2C2**	**A3B3C3**	
9028	1000	6/18	C1B2C2A3	A1B3C3	A2B1			C1B1	A3B3C3	B2A2A1C2	
9033	1000	**18/18**	C2A3B1A2B2C3C1A1B3					**B1A1C1A2C2B2A3B3C3**			
9037	1000	2/18	A3A2A1C2B2C3B1B3C1					C2B2B3C3A3C1A2A1B1			
9040	1000	3/18	B3B2A1C2A2A3C1B1C3					B3A3A1B1	B2C2C3A2C1		
9001	3000	10/18	B3A2A1	C3A3B2	C1B1C2			**A3B3C3**	**A2B2C2**	**A1B1C1**	
9003	3000	2/18	C2A1A2B2C1A3C3B1B3					**A1B1C1A2B2C2A3B3C3**			
9007	3000	13/18	B2C3C1B3A1C2A3B1A2					**A2B2C2**	**A3B3C3**	**A1B1C1**	
9018	3000	10/18	A2A3B1A1C2B3C1C3B2					Data lost			
9022	3000	10/18	A1C2A3A2B2		B3C1B1C3			B1C1A1	A2B2C2	A3B3C3	
9024	3000	6/18	A3C2A1B3C1B2C3B1A2					B3C3C2B2A3C1B1A1A2			
9030	3000	**18/18**	C3B2B3C1A1C2A2A3B1					C3B2B3A1C1C2A3A2B1			
9031	3000	**17/18**	C3A1B3C1	B1A2A3B2C2				**A2B2C2**	**A1B1C1**	**A3B3C3**	
9032	3000	15/18	A3C2	A1B3	B2C3	A2B1	C1	**A3B3C3**	**A2B2C2**	**A1B1C1**	
9039	3000	14/18	A2	B1C1C2	A1B3	C3B2	A3C2	**A1B1C1**	**A2B2C2**	**A3B3C3**	
9010	5000	3/18	C2A3B2C3B1A1B3A2C1					C2A3B2A1C1B1B3A2C3			
9012	5000	**17/18**	C1B3A2B1A3C2A1C3B2					**A3B3C3**	**C1B1A1**	**A2B2C2**	
9013	5000	11/18	C2A2C3B2A3B1C1B3A1					**C1B1A1**	**A3B3C3**	**A2B2C2**	
9025	5000	16/18	A2A3C2A1B3C1C3B2B1					**A2B2C2**	**A1B1C1**	**A3B3C3**	
9026	5000	12/18	A3A2B3B2C3C1A1B1C2					**A1B1C1**	**A2B2C2**	**A3B3C3**	
9027	5000	**18/18**	A3C2	A1B3C1	C3B2	B1A2		**A2B2C2**	**A3B3C3**	**A1B1C1**	
9029	5000	11/18	B1A3A1C1B3C2B2C3A2					**A2B2C2A3B3C3A1B1C1**			
9034	5000	**18/18**	A1B2A3	B1C1C2	A2C3B3			**A3B3C3**	**A1B1C1**	**A2B2C2**	
9036	5000	7/18	A1C1B3	A2A3C2	C3B2B1			A1B1C1A3	C2B2A2	B3C3	
9038	5000	16/18	A1C2A3C1B1A2B3C3B2					**A3B3C3**	**A2C2B2**	**A1B1C1**	

Condition A

P#	EQ	Pre Class Formation					Post Class Formation			
9009	**18/18**	A2B3B1	C1A3B2	C2C3A1			**A2B2C2**	**A3B3C3**	**B1A1C1**	
9014	6/18	C3C2C1	B2A3B1	B3A1A2			C2A2B2	A1A3	C3B3	C1B1
9015	16/18	B3A1C3B1B2A3A2C1C2					**A1B1C1A2B2C2A3B3C3**			
9016	**18/18**	B1A3A2	C3C2C1	B2B3A1			**A2B2C2**	**A1B1C1**	**A3B3C3**	
9019	7/18	B1B2A3C1C2B3A2C3A1					A2B2	A3B3C2C3	A1B1C1	
9020	**18/18**	B2B3A1	B1A3A2	C1C3C2			**A2B2C2**	**A3B3C3**	**A1B1C1**	
9028	6/18	A2B1A3	B2B3	C1C2C3A1			**A2B2C2**	**A3B3C3**	**A1B1C1**	
9033	**18/18**	A2C3A1C1B2B3C2A3B1					**A1B1C1A3B3C3A2B2C2**			
9037	**18/18**	C1C2A3A2B1C3B2B3A1					**B1A1C1A2B2C2A3B3C3**			
9040	6/18	A2A3	C3B2B3	C1C2	A1B1		**A2B2C2**	**A3B3C3**	**A1B1C1**	
9001	**18/18**	C2C3C1	A2B1A3	B2A1B3			**A1B1C1**	**A3B3C3**	**A2B2C2**	
9003	8/18	C3C2A1A2B3A3B2B1C1					C3A2B2A3B3		C2A1B1C1	
9007	**18/18**	A1C3C2C1A2A3B1B2B3					**A2B2C2**	**A1B1C1**	**A3B3C3**	
9018	**17/18**	B3B1B2A2A1A3C3C1C2					Data lost			
9022	**18/18**	A1B3C1	B2A2C3	C2B1A3			**B1A1C1**	**B3C3A3**	**B2A2C2**	
9024	15/18	A1A2B1A3B3B2C3C2C1					**A2B2C2**	**A3B3C3**	**A1B1C1**	
9030	**17/18**	A1C1A2A3B1B3B2C3C2					A1C1A2A3B1B3B2C3C2			
9031	**18/18**	A1	C2C3	C1B1	B3B2	A3A2	**A2B2C2**	**A1B1C1**	**A3B3C3**	
9032	**18/18**	A2A1	B2B1B3	C2C1A3C3			**A2B2C2**	**A3B3C3**	**A1B1C1**	
9039	**18/18**	C3C2C1	A3A2B1	B3B2A1			**A1B1C1**	**A3B3C3**	**A2B2C2**	
9010	9/18	B1A3C1C2A2A1B2B3C3					B2C1C2A2B2A1A3B1B3			
9012	**17/18**	A2B1A3	C2C3C1	B2A1B3			**B3A3C3**	**C2B2A2**	**C1B1A1**	
9013	9/18	B1A1C2B2A3C3C1B3A2					A1B1C1	B2C2	A2A3	B3C3
9025	**18/18**	B1A3C3	A1B2B3	C2C1A2			**A3B3C3**	**A2B2C2**	**A1B1C1**	
9026	**18/18**	B1C1	C3C2A1	A3B2B3A2			**A1B1C1**	**A3B3C3**	**A2B2C2**	
9027	**18/18**	B2A2B3	A3C3B1	C1A1C2			**A1B1C1**	**A3B3C3**	**A2B2C2**	
9029	**18/18**	B1B3A2B2C2A3C3A1C1					**A1B1C1A2B2C2A3B3C3**			
9034	**17/18**	C1C2C3	B2A3B3	A2B1A1			**B1A1C1**	**B2A2C2**	**A3B3C3**	
9036	**18/18**	A2B1A3	A1C2B3	C1B2C3			**A2B2C2**	**A3B3C3**	**B1A1C1**	
9038	**18/18**	A3B1	A1B2	C1C2	A2B3	C3	**B2A2C2**	**B1A1C1**	**B3A3C3**	

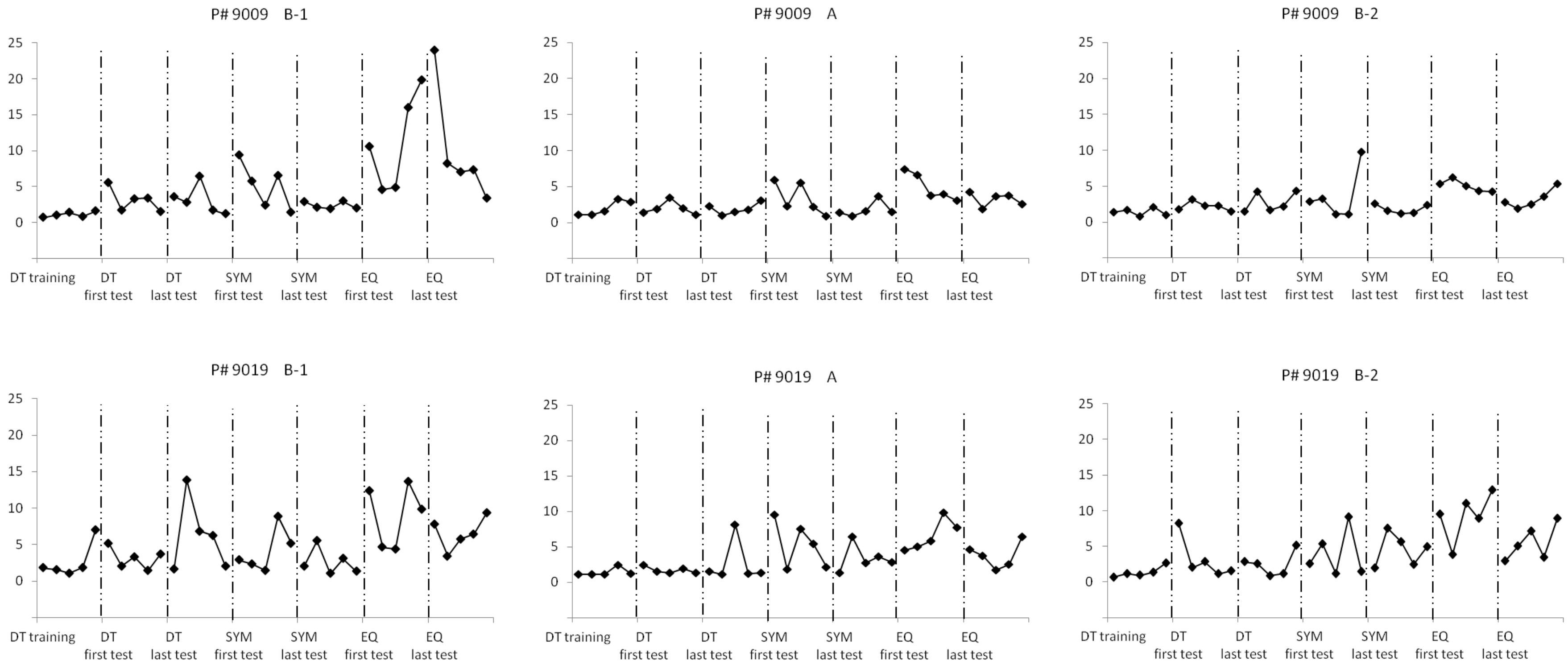
Condition B-2

P#	1N	Pre Class Formation				Post Class Formation			
9009	**18/18**	**A2B2C2**	**A3B3C3**	**A1B1C1**		**A1B1C1**	**A2B2C2**	**A3B3C3**	
9014	9/18	A3B3	A1B1C1	C3C2	A2B2	**B1C1A1**	**A3B3C3**	**A2B2C2**	
9015	**17/18**	**C1A1B1C2B2A2A3B3C3**				**A2B2C2**	**A1B1C1**	**A3B3C3**	
9016	**18/18**	**A1B1C1**	**A2B2C2**	**A3B3C3**		**C1A1B1**	**C2A2B2**	**C3A3B3**	
9019	1/18	B2C2	B3C3	B1C1	A3A2A1	B3C3	A1B1C1	A2B2C2	A3
9020	**18/18**	**B1C1A1**	**B2A2C2**	**A3B3C3**		**B2A2C2**	**B1C1A1**	**A3B3C3**	
9028	**18/18**	C1B1A1C2	B3C3A3	B2A2		B3A3C3	B2C2A1	B1C1A2	
9033	**18/18**	**B1A1C1A2B2C2A3B3C3**				**A3B3C3B1A1C1A2B2C2**			
9037	**17/18**	B2C2C3B3A3A1B1C1A2				**A3B3C3A2B2C2B1A1C1**			
9040	10/18	C3B3A3	B1A1	B2C2	C1A2	B3A3	C1B1A1	B2C2	C3A2
9001	**18/18**	A1B1C1	A3B2C2	A2B3C3		**A1B1C1**	**A2B2C2**	**A3B3C3**	
9003	16/18	**A3B3C3**	**A1B1C1**	**A2B2C2**		**A2B2C2**	**A1B1C1**	**A3B3C3**	
9007	**18/18**	**A3B3C3**	**A2B2C2**	**A1B1C1**		**A2B2C2**	**A3B3C3**	**A1B1C1**	
9018	**18/18**	C2B2A2B3C3A1B1C1A3				**A3B3C3A2B2C2C1B1A1**			
9022	**18/18**	**A3B3C3**	**A2B2C2**	**A1B1C1**		**B1C1A1**	**A2B2C2**	**A3B3C3**	
9024	**18/18**	C2B2A2	B3C3A1	B1C1A3		**A3B3C3**	**C1B1A1**	**C2B2A2**	
9030	**18/18**	B1C3C1A1A2C2A3B2B3				B1C3C1A1A2C2A3B2B3			
9031	**17/18**	**A2B2C2**	**A1B1C1**	**A3B3C3**		**A2B2C2**	**A3B3C3**	**A1B1C1**	
9032	**18/18**	**A3B3C3**	**A2B2C2**	**A1B1C1**		**A2B2C2**	**A1B1C1**	**A3C3B3**	
9039	**18/18**	**A1B1C1**	**B3C3A3**	**B2C2A2**		**A1B1C1**	**A2B2C2**	**A3B3C3**	
9010	6/18	A1B3C1B1C2C3B2A3A2				B3A1B2C2A3C3C1B1A2			
9012	**18/18**	**C3B3A3**	**C1B1A1**	**C2B2A2**		**A2B2C2**	**A3B3C3**	**A1B1C1**	
9013	12/18	C1B1A1	A3B3B2	A2C3C2		A3B3	C1B1A1	B2C3	A2B2
9025	**18/18**	**A3B3C3**	**A1B1C1**	**A2B2C2**		**A3B3C3**	**A1B1C1**	**A2B2C2**	
9026	**17/18**	**A2B2C2**	**A3B3C3**	**A1B1C1**		**A3B3C3**	**A2B2C2**	**A1B1C1**	
9027	**18/18**	**A2B2C2**	**A1B1C1**	**A3B3C3**		**A1B1C1**	**A3B3C3**	**A2B2C2**	
9029	**18/18**	**A2B2C2A3B3C3A1B1C1**				**A3B3C3A2B2C2A1B1C1**			
9034	**18/18**	**A2B2C2**	**A1B1C1**	**A3B3C3**		**A3B3C3**	**A1B1C1**	**A2B2C2**	
9036	11/18	A2B2C2	A3C1A1	B3C3B1		**A1B1C1**	**A3B3C3**	**C2B2A2**	
9038	**18/18**	**A1B1C1**	**A3B3C3**	**A2B2C2**		**A2B2C2**	**A1B1C1**	**A3B3C3**	

The figure shows responding in accordance with equivalence (EQ), and the card sorting performance for all participants prior to training and after testing in all conditions.

Figure 5. a.

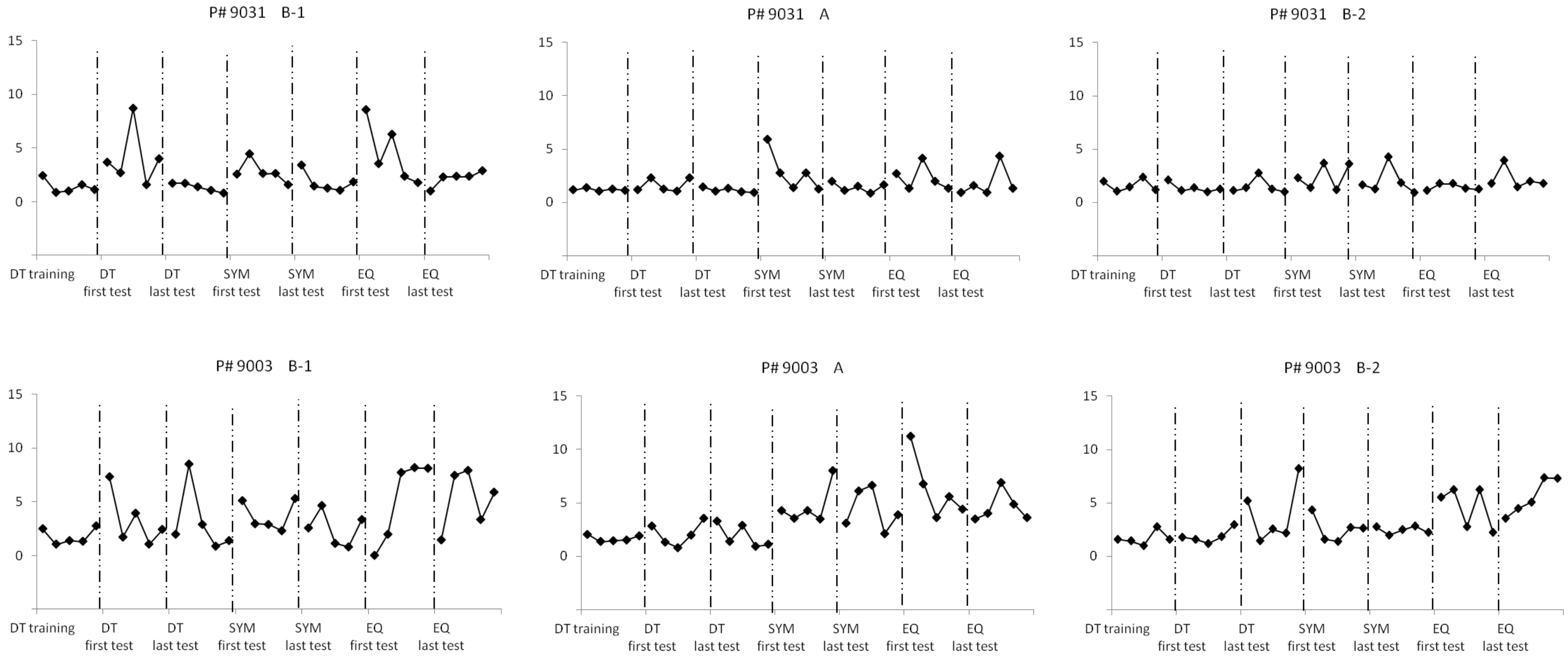
Reaction-time data for two participants in the 1000 ms group



The figure shows reaction-time data for two participants in the 1000 ms groups. The dotted lines depict the five last training trials for the directly trained relation (DT), and the five first and five last trials testing for DT relations, symmetry (SYM) and stimulus equivalence (EQ) relations. Participant 9009 responded in accordance with stimulus equivalence in all conditions, and Participant 9019 did not respond in accordance with stimulus equivalence in any of the conditions.

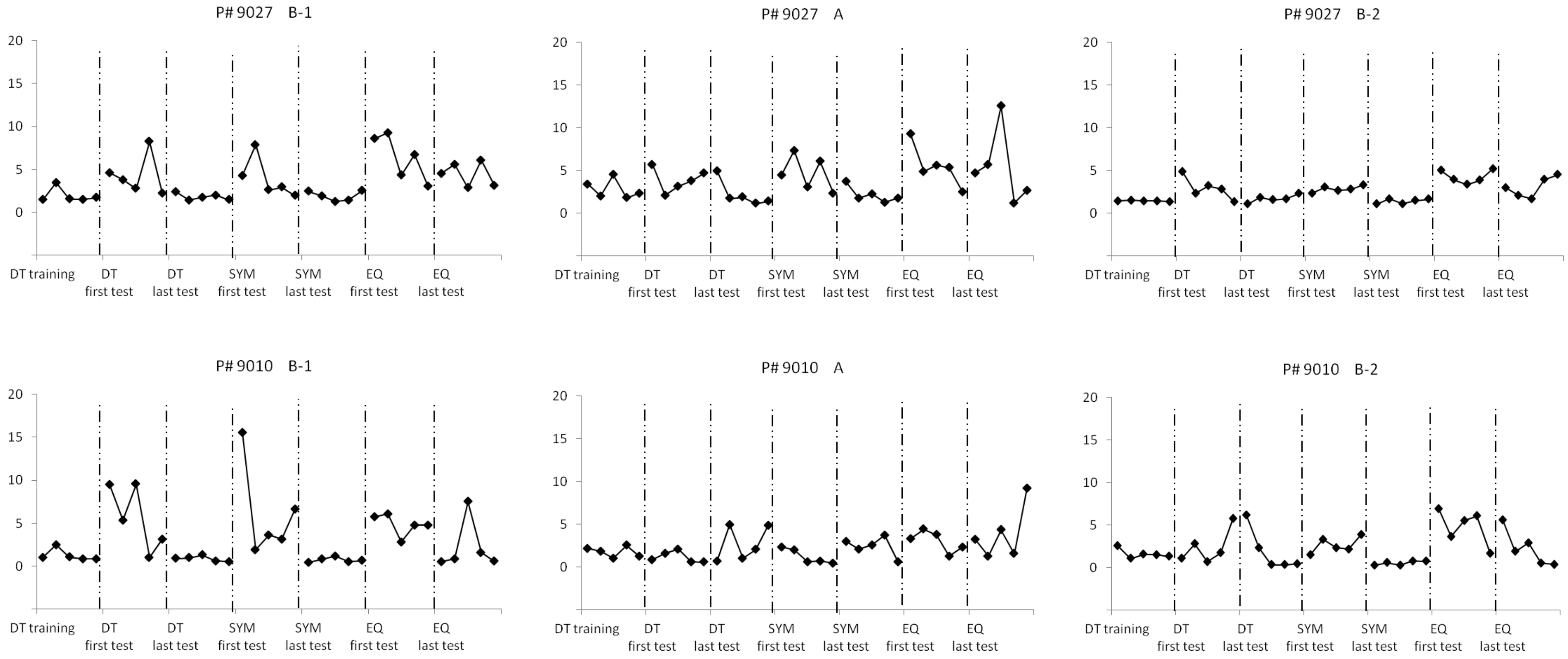
Figure 5. b.

Reaction-time data for two participants in the 3000 ms group



The figure shows reaction-time data for two participants in the 3000 ms groups. The dotted lines depict the five last training trials for the directly trained relation (DT), and the five first and five last trials testing for DT relations, symmetry (SYM) and stimulus equivalence (EQ) relations. Participant 9031 responded in accordance with stimulus equivalence in all conditions, and Participant 9003 did not respond in accordance with stimulus equivalence in any of the conditions.

Figure 5. c.
Reaction-time data for two participants in the 5000 ms group



The figure shows reaction-time data for two participants in the 5000 ms groups. The dotted lines depict the five last training trials for the directly trained relation (DT), and the five first and five last trials testing for DT relations, symmetry (SYM) and stimulus equivalence (EQ) relations. Participant 9027 responded in accordance with stimulus equivalence in all conditions, and Participant 9010 did not respond in accordance with stimulus equivalence in any of the conditions.