MASTER THESIS
Learning in Complex Systems
2013

On the Role of Sorting

Aina Dahle Norbom

Faculty of Health Sciences
Department of Behavioral Science

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Aknowledgements

I would like to thank Jon Løkke for being so helpful during the experiment and a special thanks to my supervisor Dr. Arntzen for helpful feedback and guidance during this long period. Also I would like to thank Aleksander Vie for all the technical help, and all the people that made this thesis come through.

Article 1 focuses on categorization, sorting, and stimulus equivalence. Categorization has been studied in cognitive psychology for decades and in behavior analysis the study of categorization also has played a central role through many years. Within cognitive psychology terms like categorization and concepts have been described and explained with referral to mental structures and mental processes. Within behavior analysis the terms categorization and concepts are defined as types of behavior. Sorting is a type of categorization behavior and has become a more frequent term in behavior analysis recent years. Many variables can influence results on sorting tests and therefore the present article presents some prospective influential variables on sorting results.

The experiment in Article 2 sought to find correlation between a derived relations test and a sorting test in the measures of equivalence class formation. 16 participants attended the study; three participants showed a full correlation between the two tests and responded correct on both tests, three participants sorted the stimuli in experimenter-defined classes when failing to respond in accordance with stimulus equivalence in the emergent relations test, and the remaining participants failed to form equivalence class formation in both tests. The results indicate the sorting test as an applicable additional measure of equivalence class formation.

Keywords: categorization, sorting, cognitive psychology, behavior analysis, equivalence class formation

Categorization, Sorting, and Stimulus Equivalence: Some Differences and Future Research

Aina Dahle Norbom

Oslo and Akershus University College, Department of Behavioral Science

Sorting is a process and a type of categorization behavior. In the field of cognitive psychology the term categorization and concepts are described in the name of mental structures. In the field of behavior analysis categorization and concepts are describes as a behavior and it does not refer to mental explanations. Sorting as a term will probably be more cited in the future in regards of the increasing studies using sorting in addition to other measures in research on stimulus equivalence. The present article presents different definitions of the terms categorization and concepts in regards of the two psychological views. Second it refers to studies within both fields that are current to the topic. Last, prospective future research on sorting and equivalence class formation with different types of variables, which can have effects on sorting results, is presented. Variables that will be reviewed are the role of instructions, different training structures and experimental sequence. The present article possesses the importance of the increasing use of sorting tests as additional measures in the field of behavior analysis.

Keywords: sorting, categorization, cognitive psychology, behavior analysis, equivalence class formation, instructions, training structure, experimental sequence

Categorization, Sorting, and Stimulus Equivalence: Some Differences and Future Research

The purpose of the present article is to give an overview of the terms categorization and concepts within the field of cognitive psychology and behavior analysis, and display some differences between the two psychological fields. According to Rosch (1999) who is a prototype theorist and have been cited often regarding this topic defines two basic principles in regards of categorization. The first concerns the economy of categorization and the second concerns the structure of categorization. With economy she means that it takes little effort to categorize stimuli from each other, and with structure she refers to how the world is perceived if it is structured or perceived in the way of arbitrary attributes. Thus, that categorization comes as structured as it can be, and with as little effort as possible. Keller and Schoenfeld (1950) on the other hand, wrote early about the term conceptual, that it is important not to forget it is a certain way of behaving, thus a type of behavior. They propose a question about conceptual behavior and what type of behavior it is, followed by the answer “when a group of objects get the same response, when they form a class the members of which are reacted to similar, we speak of a concept” (Keller & Schoenfeld, 1950, p. 154). Further, Keller and Schoenfeld (1950) argues that the core of concepts is stimulus generalization and stimulus discrimination. People generalize within classes, e.g., respond in the same way in the presence of different cups, and discriminate between classes, e.g., responding in one way in the presence of a glass filled with water and a tub filled with water. The basic differences between the two psychological views are the referring to mental structures in cognitive psychology, and the referring to behavior in behavior analysis in describing and explaining concepts. Both fields need additional experimentation of concept formation, and the present article describes some different theories and studies within both fields. Further, the current standpoint of equivalence class formation in regards of sorting tests within stimulus equivalence research will be reviewed. The article will give a pinpoint on future research regarding sorting tests as an additional measure in the field of behavior analysis and stimulus equivalence. Additionally, different procedures that could take place in regards of the sorting test as an additional measure on equivalence class formation. The following different variables are predicted to have prospective effects on the sorting results which could serve as a reference for future research on this topic. The different variables that will be presented in this article is the role of instructions, different sorting procedures, and training structures, how a sorting test could be used as an additional measure to compare different procedures in a vast and time effective way, and different experimental conditions.

Various Definitions of Categorization: Cognitive Psychology

Categorization has been cited and studied through many years within the psychology field. Especially in the field of cognitive psychology (Ashby & Maddox, 2005; Barsalou, 1991, 1992; Eysenck & Keane, 1995; Margolis & Laurence, 1999; Rosch, 1975, 1999; Smyth, Collins, Morris, & Levy, 1994) the term categorization have and still provides a major contribution to the learning of concept formation. Margolis and Laurence (1999) edited a book containing different papers concerning the terms concept and categorization. The book consists of descriptions of five different theories of concepts and categorization. The first theory is The Classical Theory and is defined the basis for other theories in the sense that it is dated back to Plato. The Classical Theory's basis is that "most concepts (esp. lexical concepts) are structured mental representations that encode a set of necessary and sufficient conditions for their application, if possible, in sensory or perceptual terms" (p. 10). The theory differentiates between what defines concepts and what does not e.g., a telephone. A basic telephone is characterized by numbers to dial, a speaker and a microphone. If one of those does not exist, say the speakers, it is not characterized as a telephone. All the characteristics need be present to define the concept of telephones. A model of categorization is also offered in The Classical Theory as an extension of the theory. Categorization in regards of this theory states that features of an item must fall under the category being categorized to a certain concept. Using the telephone example again, can the numbers on the phone be categorized under the concept of phones or must they contain other characteristics. The theory faces several problems and that is a reason for development of new theories. The second theory is The Prototype Theory. This theory is alternative to The Classical Theory and forced its downfall. In the case of the first theory problems can arise when not all characteristics of a concept are present, but the concept cannot be dismissed nevertheless. If the telephone had an antenna, that would postulate difficulties. New telephones, compared to old ones are not equipped with antennas, but still get categorized under the concept of a telephone. The Prototype Theory states that not every feature is necessary to form the concept. A problem with this theory is the way it is defined. The theory forces concepts as difficult to apply in the wrong way, thus, they “play it safe”. Consider the example of a blueberry, it is blue, the size of a pea, the plant is green and it is dark inside. Next to the blueberries another type of berry grows, which is blue, on the size of a pea and the plant is green. It really looks like a blueberry, except it is not. The color on the plant is lighter and the leaves are bigger and the blue color on the berry is lighter than the color of the blueberry. Another problem is that the prototypes are set, thus, a class has specific features, and if some of these features are not present it is no longer a typical glass. This is defined as more typical and less typical, so the fewer features of each item, the less typical of that concept (Laurence & Margolis, 1999). The problem is that prototype concepts as e.g., even numbers are differentiated as being more or less typical, that 8 is more prominent as an even number than 34 (Armstrong, Gleitman & Gleitman, 1983, in Laurence & Margolis, 1999). Laurence and Margolis (1999) further states that prototypes generates a problem in that a concept is not always prototypical, only a few people can categorize specific features that forms a concept which other people are not acquainted with e.g., Mr. Anderson, your neighbor who is grumpy, alone, grey hair, has a dog and drives a Honda. People's neighbors are not prototypical for everyone and that makes it difficult to define prototype concepts because there are so many non-prototypical concepts.

The third theory is The Theory-Theory of Concepts. The definition of this theory is that "concepts are representations whose structure consists in their relation to other concepts as specified by a mental theory" (Laurence & Margolis, 1999, p. 47). The theory relies on mental structures of a category and the theory of that category is individualized by the body of knowledge. The theory in which the concept is formed is of interest, and the mental representation of one concept is in contrast to mental representation of another concept. In difference to the before mentioned theories, this theory does not rule out a concept in the lack of characteristics. The view of this theory is psychological essentialism meaning that objects' hidden properties or internal structure decides which category is appropriate, and not the mere physical properties of the object. For example take the concepts of a house. The physical properties of a house contain four walls, windows, a roof, one or more doors etc. A cabin has almost the same features as a house, but in the presence of a cabin, most people categorize the smaller house in the category of a cabin. Since it is small it could also be categorized in the category of a hovel, but people still differentiate between these concepts even though the objects have the same properties of a house. Thus, when making category decisions people access a mentally represented theory. The hidden property is essential in decision making of categorizing of the object. An obvious problem is the notion that people have problems explaining properties that differentiates one object from another. Take the house example again; it would be hard to differentiate the properties that make a house and the properties that make a cabin.

Moving forward to the fourth theory of concepts is The Neoclassical Theory of Concepts, and it expands on The Classical Theory. The difference is that neoclassical theorists are more sensitive to problems mentioned in regards of The Classical Theory (Laurence & Margolis, 1999). Specially linguists have taken an interest in this theory and find the meaning of the words as important and that internal categorization is present when forming a concept (Jackendoff, 1983). The theory does not actually contain any specific domain; rather it is a collection of psychological standpoints that share common features and explanations. Jackendoff (1983), a neoclassical theorist described four characteristics of categorization that applied to verbal (labeling) and nonverbal (discrimination) categorization. First, a set of rules must control the judgments. Second, these specific rules tend to be unconscious. Third, the pattern of judgments is constructed of yes/no/not sure. Fourth, he argues that borderline cases, the truth of judgments is odd to speak about. He also states that information and processes is of importance when the judgments of the categorization are present.

The fifth and last theory of concepts is the Conceptual Atomism theory. This theory goes against all the previously described theories because it dismisses the assumption that all concepts have structures. Thus, this theory does not apply their explanations of concepts in the mental structures; rather they argue that concepts have no structure. An example taken from Laurence and Margolis (1999) is that concepts of bird is not categorized in the features of wings, feathers etc., it is the concepts of bird itself. Other casual laws determine a bird as a bird without applying to other physical features or mental structures. The theory states that lexical concepts are primitive. There are some problems to this theory, one of them being the problem stating that concepts are innate, if not learned they must be innate. How are they going to explain the concepts that are innate because they are not learned? How is it then possible to categorize into a concept if it is not learned? Few people have committed to this theory (Laurence & Margolis, 1999).

Palmer (2002) reviewed the book of Margolis and Laurence (1999) argues in regards of the definitions “the term *concept* is seen by some of these authors as a natural phenomenon whose properties must be discovered rather than a formal category whose properties can be prescribed by definition" (p. 598). He later points out that most authors defines concepts as things that can be studied and not be in need of a definition because the term is too familiar. It seems the terms concept and categorization is difficult to define in simple matters and with principles of selection. Rosch (1999) describes categorization as having a vertical and horizontal dimension, meaning the vertical dimension is the distinction concerning the inclusiveness of the category, e.g., the distinction between mammals, dogs, different kinds of dogs, living things etc. The horizontal dimension refers to the inclusiveness concerning the segmentation of the category, e.g., discrimination between a cow and a horse, a mouse and a hamster and so forth. This resembles the processes referred to in behavior analysis literature as generalization and discrimination, but the difference being that behavior analysis does not refer to dimensions beyond our analysis.

Ashby and Maddox (2005) focused on perceptual categorization and how people learn this, and made a distinction between perceptual categories and concepts. The distinction lies in the product meaning when related thing are part of a group we talk about perceptual categories, and when different ideas together form something e.g., a religion, we talk about concepts. In addition they stress four sub literatures that will only be mentioned briefly. The first concerning non-humans, the second concerns highly experienced experts which they do not focus on, the third concerning focus on how people learn new categories, and the fourth concerning their focus on perceptual categories rather than concepts. The term perceptual categories or perceptual classes has been described and defined in different ways than by Ashby and Maddox (2005), but still project some similarities. Fields et al. (2002) proposed that three criteria has to be present before a perceptual class has emerged. First, all stimuli in a specific class have to be followed by the same response. The important aspect of this is that only some of the stimuli have had direct reinforcing contingencies, but still every stimuli occasion the same response. Second, there has to be a higher probability of the response in the presence of stimuli that has anteceded the response, and vice versa, that is, a lower probability of the response in the presence of the stimuli that is unreinforced. Last, the domain stimuli represents must be discriminated from each other. Fields et al. (2002) also describe the term generalized categorization which means that people categorize domains never before trained, and do this spontaneous. They argue this as a reason for several deficits in regards of the impediment of learning in people with dementia, autism or mental retardation.

Smyth et al. (1994) defined categorization as how people group things together. He also defined the concept of abstraction, which includes how people make distinctions between those things (e.g. a sheep and a goat). Eysenck and Keane (1995) wrote about the knowledge that humans through genetics are born with, and that this knowledge is learned trough experience. The key was that knowledge needs to be organized in an economic and informative fashion. They mean that “we generalize from instances of objects by noting their similarities and ruling out their differences” (p. 235). As mentioned before these kind of descriptions of concepts and categorization is much similar to the definitions of generalization and discrimination in the field of behavior analysis. Keller and Schoenfeld (1950) defined stimulus generalization in the way that in the presence of certain stimuli, people respond in the same way. Stimulus discrimination was described as developed in the face of generalization and that this discrimination refers to responding in the presence of a certain stimulus and not in the presence of another.

Many similarities can be found between cognitive psychology and behavior analysis when looking thorough at the variables described in the definitions of categorization and concepts. The differences lie especially in explaining the variables and how they affect certain behavior. The biggest difference being that cognitive psychologists explain the behavior of organisms in terms of mental structures while behavior analysts concentrate on the behavior itself and what can be observed.

Various definitions of categorization: Behavior analysis

Similar to cognitive psychology the field of behavior analysis also has focused on categorization for decades. There is no opposition to concept and categorization as complex human behavior. There are no easy descriptions of what happens when categorization takes place and allegedly concepts are formed. According to Donahoe and Palmer (2004) the term concepts is an nontechnical term meaning that it arises from everyday language, “they are not the product of a consistent set of basic processes” (p. 129). In this regard a category is theoretically incoherent. Another term that should be applied instead of concepts in behavior analytic literature is stimulus classes. Selection history and the guiding of the stimulus are of high value and if constructs are not sensitive to this, then it is theoretically incoherent. People form concepts around different things e.g., a plane or a cat. The problem in referring to these as concepts is that similar variables affect the “concepts” differently when it in general is considered to be the same things. These stimulus classes as just described are called discriminative stimulus classes and is a type of stimulus class among others. What defines a discriminative stimulus class is the different stimuli in the same class consists of physical similarities (Donahoe & Palmer, 2004). Other types of stimulus classes are functional classes. Functional classes are defined by their functional similarity so they do not need physical similarities. The key is that the classes control common responses. For example if one says “flying object” in the presence of a commercial airplane, the response “flying object” would also be present in the presence of a helicopter. The two stimuli are both flying object, but lack physical similarities. Still they belong to the same functional stimulus class (Donahoe & Palmer, 2004). Miguel and Petursdottir (2009) proposes a simplified definitions of categorization and concepts meaning that categorization (or classification) is when objects are grouped together and a category (or class) consists of the coherence of these objects.

Sidman (1994) used the term partition as another term for classification or categorization, meaning that things are separated e.g., hats, shoes etc. If partition is possible then "pairs of components within each class are included in an equivalence relation" (p. 417). Some examples are described where red object, green object and blue objects are partitioned in regards of their colors and therefore green objects are equivalent to each other, similar to blue and red objects. He argues that categorization is possible with objects that are similar, physically different, and common function is what classifies the stimuli that are physically different. He concludes that equivalence relations and partition (or categorization) is the same, but definitions and testing is two different things. It is not the same behavior when categorizing or responding in regards of conditional discrimination as responding in accordance with stimulus equivalence (Sidman, 1994). He also states that partitioning is analogous to functional classes, that instead of using the term partitioning the term functional class can be applied since it is a behavioral name for partition class (Sidman, 1994; Sidman, Wynne, Maguire, & Barnes, 1989). The background for this statement derives from Sidman et al. (1989). On the basis of the results from Vaughan (1988) showing that a different discrimination procedure was used to establish concept formation, Sidman et al. (1989) found results indicating that functional classes and equivalence relations are the same. A big difference between the two studies was the use of pigeons (Vaughan, 1988) and the use of humans (Sidman et al., 1989). Two out of three participants in Sidman et al. (1989) showed functional classes related by equivalence. Even though functional classes and equivalence relations are reconsolidated, differences in definitions and testing procedures are present, but Sidman (1994) sees it as an advantage the differences in definitions and testing with the statement "the less plausible a prediction, the more significant its confirmation" (p. 447).

With this statement, use of sorting or categorization tests in equivalence research should be more prominent.

When performing studies in regards of categorization it is variables with effect on categorization that is of interest. In stimulus equivalence research there has been a small growth in studying categorization. The term categorization will be described in regards of how it is defined in stimulus equivalence and behavior analysis, as a kind of behavior and not a mediating process facilitating mental processes. Researchers in the field of behavior analysis and stimulus equivalence has also conducted several studies regarding categorization (Herrnstein & Loveland, 1964; Horne, Lowe, & Harris, 2007; Horne, Lowe, & Randle, 2004; Lowe, Horne, Harris, & Randle, 2002; Lowe, Horne, & Hughes, 2005; Miguel & Kobari-Wright, 2013; Miguel & Petursdottir, 2009; Miguel, Petursdottir, Carr, & Michael, 2008; Sigurdardottir, Mackay, & Green, 2012; Zentall, Galizio, & Critchfield, 2002), most of all concerning the emergence of categorization through different types of training structures. Miguel and Kobari-Wright (2013) explored categorization and if naming produces listener behavior and categorization, and whether naming is necessary for categorization to occur. The categorization test was conducted by presenting a picture from the categories and then participants were told to point at one comparison (pretests) and asked “what is this”? (posttests). The results showed that tact training alone can be used in the emergence of categorization to occur. They point to the fact that the study only had two participants and that future research should replicate the experiment. However, the important variable in regards of the present article is the use of procedures to test categorization.

In recent years the term sorting has been cited more often. The sorting test is a test where participants group stimuli together, both with laminated cards and also on computers. Common for many of the studies that use sorting is the use of these procedures as additional measures presented as a pretest and posttest, but the sorting seldom is the main focus of experiments. Sigurdardottir et al. (2012) conducted a study of stimulus equivalence, generalization and contextual stimulus control in verbal classes. Among the variables studied, they used a sorting task to evaluate the transfer of contextual control in Experiment 2. Fields, Arntzen, and Moksness (2013) used sorting as an additional index on equivalence class formation and found that participants responded in accordance with stimulus equivalence also sorted stimuli in experimenter-defined classes in the sorting test. They also found that different classes formed by participants during the MTS-test also were sorted together in the sorting test. For example, if a participant had established class 2 in the MTS-test, then this class also was sorted together. The results of the experiment is positive in the argumentation of sorting tests to measure equivalence class formation responding in addition to the standard measure in the field of behavior analysis. To emphasize the terms used in the present article, when referring to stimulus equivalence it means equivalence classes participants have established through MTS procedures. When referring to equivalence class formation it means experimenter-defined classes that participants have sorted together in a sorting test. Thus, when they sorted stimuli into experimenter-defined classes then participants have established equivalence class formation. If they sort the stimuli in participant-defined classes then the participants has not responded in accordance with equivalence class formation. Likewise, if participants do not respond within the criteria of the MTS-test, then participants have failed to respond in accordance with stimulus equivalence. Further, in the present article the term sorting will be used and not categorization. The term is a type of categorization in the way that it is possible to measure how participants categorize stimuli through the sorting test. The following paragraphs will contain future prospective research regarding the sorting test and variables that can influence the results that should be experimented further.

Instructions

Within experimental behavior analysis and stimulus equivalence, studies have been conducted regarding instructions and the effects instructions have on MTS-tests (Devany, Hayes, & Nelson, 1986; Green, Sigurdardottir, & Saunders, 1991). Devany et al. (1986) compared language-able and language-disable children and found a correlation between language-able children and language-disable children and the results of tests for stimulus equivalence. The language-able children formed equivalence classes, but in contrast the language-disable children failed to form equivalence classes. These results indicate that instructions play a central role in the establishment of stimulus equivalence. Thus, if participants do not get instructed before training and testing, they would probably not respond in accordance with stimulus equivalence on the MTS-test. Similar, if participants get instructed they most likely will form equivalence classes. There has also been conducted studies in experimental psychology that has the main focus on instructions in regards of various tasks not involving stimulus equivalence (Cabello, Luciano, Gomez, & Barnes-Holmes, 2004; Hayes, Brownstein, Haas, & Greenway, 1986). In regards of the results from the Devany et al. (1986) the use of instructions is crucial in the establishment of stimulus equivalence, and various instructions can be applied and is an important aspect. Sidman (1992) stressed the importance of the type of instructions that are presented to participants. Whether the rules give rise to equivalence or if equivalence makes rules possible needs to be further examined. If participants get instructions stating that stimuli “go together”, which often are the case in these studies, the results could be positive in regards of responding in accordance with stimulus equivalence. The responding in accordance with stimulus equivalence is probably not due to the experimental conditions, but due to rules of the experiment described in the instruction. Sidman (1992) states this by saying that instructions alone can have control over participants responding and that the experimental conditions exerts no control over the responses.

Other studies have examined instructions as a variable with important effects on behavior (Drake & Wilson, 2008; Green et al., 1991). Drake and Wilson (2008) studied the effects of instructions on MTS-responding with students. 11 participants received additional instructions compared to the remaining 20 who received a standard instruction, which every participant received when starting the experiment. The additional instructions enclosed two different rules, the first stating they would get 1.5 hours of credit independent on time spend on the experiment, and the second stating that number of correct responses equals finishing more quickly. The results showed that participants receiving additional instructions had more correct responding and higher levels of completion on tests compared to participants who did not. However, in contrast to Drake and Wilson (2008) a different argument was made by Green et al. (1991) in regards of instructions as an effect changing variable. They examined whether a minimum of instructions would have different effects on participants responding than participants who did not have instructions at a minimum. Eight participants attended the study, four of them were allocated to a control condition, three were allocated to the standard instruction and one participant was allocated to the condition with a minimum of instructions. The results showed that participants in the control group did not transfer functions, the three participants in the standard instruction group did transfer functions along with the participant who had minimally instruction. These results indicate that there is no need for much instruction to mediate responding. However as they point out as a limitation only one participant received minimal instructions and one participant alone is not enough to conclude that instructions are not needed to find positive responding in regards of the behavior under study.

Another kind of instructions applied in experiments is self-instruction by participants, in which the participants performs as both listener and speaker (Baron & Perone, 1998). Rosenfarb, Newland, Brannon, and Howey (1992) conducted a study concerning the effects on self-generated rules. Participants were allocated into three different groups; the first group was instructed to make rules during the experiment, the second group was yoked to the first group and also had to make rules but were not formally instructed to do so, instead behavior was controlled by direct contingencies. The third group had no additional instructions and was not presented with any rules during the experiment. The results showed that groups receiving rules responded to the contingencies of the experiment faster than the group without further instruction. An interesting result was during extinction, the no-rules group changed their responses faster in line with the changes in experimental contingencies. This is in contrast to the other groups that made rules. The rule-groups showed these changes at a lower rate in the extinction phase. These results indicate that behavior is harder to change when under control of complex contingencies rather than direct experimental contingencies. Additionally, in the sorting literature the application of instructions are found, and in the literature regarding stimulus equivalence there has been conducted some studies that has used sorting as an additional measure on equivalence class formation (Eikeseth, Rosales-Ruiz, Duarte, & Baer, 1997; Eilifsen & Arntzen, 2011; Fields et al., 2013; Fields, Arntzen, Nartey, & Eilifsen, 2012; Smeets, Dymond, & Barnes-Holmes, 2000) and it is a growing topic in this line of research. Most of these studies have used laminated cards and some of them have used paper-and-pencil formats to measure stimulus equivalence responding and equivalence class formation in the form of sorting tests. Typical for sorting tests is that participants are asked to put stimuli into groups followed by a specific procedure for that specific sorting task (e.g. paper-and-pencil or laminated cards). The instructions for the sorting tasks could be presented in various ways and as mentioned before, the type of instructions can be crucial for the results. For future research, instructions for the sorting test should be combined with the instructions for the MTS procedure to make sure participant hold the sorting response and MTS-response under the same instructional control. If participant does not, it could make the participants results not correlate with the results from the sorting test and the MTS-tests. This could be because instructions from the sorting test states one thing and the instructions from the MTS procedure stated another, which in turn could make the participant continue to sort the stimuli in a participant-defined classes through the experiment and show no correlation between the sorting test and the MTS-test. Additionally, sorting test entails a whole different procedure then the MTS procedure so it can be difficult for participants to understand how these two tests correlate. See previously mentioned studies for descriptions of different sorting procedures. It is important to remember that a sorting test is only an additional test. It is not a procedure for establishing equivalence class formation; rather it is a process that examine if participant have responded in accordance with equivalence class formation.

Another way instructions can interfere with the results is for obvious reasons; what exactly is said to the participant. What do the instructions state to the participants? Instructing participants to sort stimuli together could influence the participant to instantly put the stimuli into groups with similar shape or form when applying words like “sort” or “put together”. This brings us back to the problem pointed out by Sidman (1992) about phrases that can lead participants to form rules that control the sorting response. The sorting response is the response that participants perform in the presence of a presented sorting test. The response is when participants move stimuli around and put these into groups that are either defined by the participant him- or herself, or into groups that are consistent with the classes defined as correct by the experimenter. An important aspect of a sorting response is that it is not a sorting response before participants actually sort the stimuli into groups. Again, instructions are an important variable of the sorting. If participants do not get instructions to group stimuli together, they probably would not display a sorting response. When instructing participants to sort stimuli together as “you feel like”, the critical part is “as you feel”. If participants had received instruction to “sort them together”, maybe it would give an indication about groups with the same amount of stimuli in each, that in the presence of 15 different stimuli there should be either three groups containing five stimuli or five groups containing three stimuli. When adding “feel like”, the sorting response is most likely controlled by learning history in regards of the term “free” and not “sort them together”. It does not give an indication of a predetermined answer to how the stimuli should be sorted. The learning history of participants plays a central role in how the instructions affect responding. When people undergo experiments they have no knowledge about, they often believe they are being tested for their IQ or that what they do can expose their personality etc. “Feel like” would not interfere with the experimental conditions or give information about the procedures or purpose of the study. In the pretest it is important that no instruction indicating experimental procedures are presented. However, in the posttest a higher probability is that the sorting response is controlled by experimental conditions from the MTS-training and testing rather than the instruction “sort together”.

Training structures

In research on stimulus equivalence there are three different training structures that are mostly described; many-to-one (MTO) training structure, one-to-many (OTM) training structure and a linear series (LS) training structure (Arntzen, 2010). The topic of training structures has been central in several studies comparing OTM, MTO and LS (Arntzen & Hansen, 2011; Arntzen & Holth, 1997), and comparing only OTM and MTO (Arntzen, Halstadtro, Bjerke, & Halstadtro, 2010; Arntzen & Nikolaisen, 2011) The studies indicate training structures as profound in establishing equivalence classes. Most studies have shown LS training structure as more difficult in establishing stimulus equivalence when comparing OTM or MTO with the LS training structure. In regards of previous stimulus sorting studies some have used MTO (Arntzen, 2004), OTM (Galizio, Stewart, & Pilgrim, 2001; Grimm, 2011), and LS training structures to establish stimulus equivalence (Arntzen, Braaten, Lian, &

Eilifsen, 2011; Smeets et al., 2000). An interesting research question concerning training structures and the outcome on the test for derived relations is prospective difference on the sorting tests compared to different training structures. It could be assumed that sorting tests yields higher correlations between the sorting test and the tests for derived relations when trained with an MTO or OTM training structure compared to the LS structure. This because previous studies have shown when participants responds in accordance with stimulus equivalence they also sort the stimuli in experimenter-defined classes on the sorting test (Arntzen, 2004). Comparing of these structures should be conducted to measure prospective differences in the outcomes of sorting tests using a MTO, OTM, or LS training structure. Instead of conducting tests for derived relations, a sorting test alone could be applied to measure if equivalence classes are established. The basis for this assumption, as previously discussed in this article in the light of the discussion by Sidman (1994), the test for derived relations could be necessary for the establishment of equivalence classes.

If one want to study different training structures it should be done with several participants. This of course takes a lot of time as shown in most studies on stimulus equivalence. Each participant spends some time to conduct a test for derived relations and the training itself takes a long time to conduct, and the length depends on the number of classes and number of members in those classes. Suppose sorting tests could be applied as an additional measure, then experiments could be conducted in less time and more experiments in a shorter time could be conducted. When applying only the sorting test, which structure that yields higher results on equivalence class formation can be measured quickly, and if the test for derived relations really is necessary for the establishment of stimulus equivalence.

Applying both the test for derived relations and sorting test when comparing different training structures should be conducted, and in addition applying a sorting test to measure the correlation between the two tests. It can be assumed that results on the sorting test should derive from the different training structures. Thus, introducing a sorting test right after the training procedure and comparing the same procedure with different training structures, MTO and OTM should yield sorting responses with better results than with participants presented with a LS training structure. Presenting the sorting test right after the training procedure would give the experiment higher control because the test for derived relations has not mediated the establishment of the equivalence classes. Therefore, a sorting test can give a more precise measure of equivalence class formation with each participant. If results show that no participant sort the stimuli in experimenter-defined classes right after the test, then that indicates that the tests for derived relations could be a necessity in the establishment of equivalence class formation. If participants sort the stimuli in experimenter-defined classes right after training that yields evidence for another positive side to the sorting test as an additional measure.

Experimental sequence

Sorting tests conducted in previous research has applied these as pretests and posttest (Arntzen et al., 2011; Fields et al., 2012; Smeets et al., 2000). This kind of procedures gives a good indication on sorting as an additional measurement of equivalence class formation. However, variations of the sequences of the experiment should be conducted to show a greater indication of the sorting test as an additional measurement. Instead of using only sorting tests as pretests and posttests it can be intervened in other variations throughout the experiment. Intervening sorting tests between training and testing for derived relations, use of repeated sorting tests after training, sorting tests after a repeated number of tests for derived relations etc., are possible variations of the presentation of sorting tests.

A possible limitation of sorting tests intervened between tests for derived relations and training procedures can be that sorting tests can facilitate improved results on tests for derived relations. Thus, if sorting tests is presented before tests for derived relations the question of whether the sorting test affected the derived relations test need answering. Did the presentation of the sorting test make it easier for participants to form classes in the derived relations test? Threats to internal validity has been cited thoroughly in the literature (Catania, 2007; Cooper, Heron, & Heward, 2007) and presentations of sequences is one threat. When presenting several variables, the first can have effect on the second. Sorting tests for example, is itself only a test for equivalence class formation. If one presents this test on different times during the experiment the sorting test could have an effect on the results of the MTS-test. Thus, participants have a greater chance at failing the MTS-test if the sorting test is not presented. This should be examined further through experimentations by presenting one group with a sorting test after training and one group with a sorting test after the test for derived relations and compare the results. Suppose the results of the groups do not display any difference on the MTS-tests, then sorting tests prove to be more applicable in equivalence research.

To make sorting tests even more applicable the number of reintroductions of sorting tests and MTS-tests should be applied. This is a matter of delayed emergence. Would participants change their sorting response after repeated introductions to sorting tests? The question can be answered if reintroduction of tests is applied and participants change their sorting response to more experimenter-defined classes in the second sorting test than in the first test. If these results were to be found, that reintroductions of the sorting test alone would produce higher yields of equivalence class formation for each participant, then it could be assumed that sorting tests can mediate some kind of response in the participants repertoire. Sidman (1994) pointed out the area of delayed emergence, that tests for derived relations could be arbitrative for the participants to establish equivalence relations. Thus, training procedures alone won’t make participants establish classes. Studies have shown that participants, who do not respond in accordance with equivalence on the first test for derived relations, do so when the test is reinstated. The test presents no feedback, dismissing that the participants prospective better responding derive from experimental training procedures. In this example, tests for derived relations is presented right after training, but suppose the test is not introduced at all, only sorting tests. Interesting results would show more experimenter-defined classes after further presentations of the sorting tests.

Another way to measure if sorting tests have any effect on results of equivalence responding is to present the sorting test after the MTS-test and not reinstate it before numerous presentations of the MTS-test is conducted after the sorting test. Participants who do not respond in accordance with stimulus equivalence or sort stimuli in experimenter-defined classes must be exposed to these conditions. If participants does not respond in accordance with stimulus equivalence after several MTS-tests has been conducted, then a sorting test do not have effect on the responding on the test for derived relations and can be used as an additional measure of equivalence class formation. Also for those participants that are sorting stimuli in experimenter-defined classes but still failing to respond in accordance with stimulus equivalence makes sorting tests more applicable. Moreover, replications of present studies on the topic of experimental sequences should be conducted in a greater amount since the literature is quite narrow at the current moment.

Conclusion

Categorization and concepts are describes within the field of cognitive psychology and behavior analysis, but display great differences in the description and explanation of what is occurring when someone categorize objects that make a concept. Sorting as a type of categorization behavior is a topic in the need of further experimentation. Sorting also need further research in other parts of behavior analysis to prove its prospective applicability. Still, there are many variations of this procedure that can be emitted to find different results, and should be adopted by scientists in the future.

References

Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *The Psychological Record, 54*, 275–291. Retrieved from http://opensiuc.lib.siu.edu/.

Arntzen, E. (2010). Om stimulusekvivalens. In S. Eikeseth & F. Svartdal (Eds.), *Anvendt atferdsanalyse: Teori og praksis* (2 ed., pp. 100–138). Oslo: Gyldendal Akademisk.

Arntzen, E., Braaten, L. F., Lian, T., & Eilifsen, C. (2011). Response-to-sample requirements in conditional discrimination procedures. *European Journal of Behavior Analysis, 12*, 505–522. Retrieved from https://oda.hio.no.

Arntzen, E., Halstadtro, L. B., Bjerke, E., & Halstadtro, M. (2010). Training and testing misic skills in a boy with autism using a matching-to-sample format. *Behavioral Interventions, 25*, 129–143. doi: 10.1002/bin.301

Arntzen, E., & Hansen, S. (2011). Training structures and the formation of equivalence classes. *European Journal of Behavior Analysis*, *12*, 483–503. Retrieved from https://oda.hio.no.

Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record, 47*, 309–320. Retrieved from http://opensiuc.lib.siu.edu/.

Arntzen, E., & Nikolaisen, S. L. (2011). Establishing equivalence classes in children using familiar abstract stimuli and many-to-one and one-to-many training structures. *European Journal of Behavior Analysis, 12*, 105–120. Retrieved from https://oda.hio.no.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual review of psychology, 56*, 149–178. doi: 10.1146/annurev.psych.56.091103.070217

Baron, A., & Perone, M. (1998). Experimental design and analysis in the laboratory study of human operant behavior. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 45–91). New York: Plenum.

Barsalou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The Psycology of Learning and Motivation* (Vol. 27, pp. 1–64). San Diego: Academic Press, Inc.

Barsalou, L. W. (1992). *Cognitive psychology: An overview for cognitive scientists*. Hillsdale, New Jersey: Lawrence Erlbaum Associates

Cabello, F., Luciano, C., Gomez, I., & Barnes-Holmes, D. (2004). Human schedule performance, protocol analysis, and the "silent dog" methodology *The Psychological Record, 54*, 405–422. Retrieved from http://opensiuc.lib.siu.edu/.

Catania, A., C. (2007). *Learning* (4 ed.). Cornwall-on-Hudson, NY: Sloan Publishing.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied Behavior Analysis* (2 ed.). Upper Saddle River, New Jersey: Pearson Education, Inc.

Devany, J. M., Hayes, S. C., & Nelson, R. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior, 46*, 243–257. doi: 10.1901/jeab.1986.46-243

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

Drake, C. E., & Wilson, K. G. (2008). Instructional effects on performance in a matching-to-sample study. *Journal of the Experimental Analysis of Behavior, 89*, 333–340. doi: 10.1901/jeab.2008.89-333

Eikeseth, S., Rosales-Ruiz, J., Duarte, A., & Baer, D. M. (1997). The quick development of equivalence classes in a paper-and-pencil format trough written instruction. *The Psychological Record, 47*, 275–284. Retrieved from http://opensiuc.lib.siu.edu/.

Eilifsen, C., & Arntzen, E. (2011). Single-subject withdrawal designs in delayed matching-to-sample procedures. *European Journal of Behavior Analysis, 12*, 157–172. Retrieved from https://oda.hio.no.

Eysenck, M. W., & Keane, M. T. (1995). *Cognitive psychology: A student's handbook* (3 ed.). UK: Psychology Press Ltd.

Fields, L., Arntzen, E., & Moksness, M. (2013). Stimulus sorting: A quick and sensitive index of equivalence class formation.

Fields, L., Arntzen, E., Nartey, R. K., & Eilifsen, C. (2012). Effects of a meaningful, a discriminative, and a meaningless stimulus equivalence class formation. *Journal of the Experimental Analysis of Behavior, 97*, 163–181. doi: 10.1901/jeab.2012.97-163

Fields, L., Reeve, K. F., Matneja, P., Varelas, A., Belanich, J., Fitzer, A. et al. (2002). The formation of a generalized categorization repertoire: Effect of training with multiple domains, samples, and comparisons. *Journal of the Experimental Analysis of Behavior, 78*, 291–313. doi: 10.1901/jeab.2002.78-291

Galizio, M., Stewart, K. L., & Pilgrim, C. (2001). Clustering in artificial categories: An equivalence analysis. *Psychonomic Bulletin & Review, 8*, 609–614. doi: 10.3758/BF03196197

Green, G., Sigurdardottir, Z. G., & Saunders, R. R. (1991). The role of instruction in the transfer of ordinal functions through equivalence classes. *Journal of the Experimental Analysis of Behavior, 55*, 287–304. doi: 10.1901/jeab.1991.55-287

Grimm, L. J. (2011). *Emergence of equivalence relations: comparing sorting and match-to-sample procedures.* Nortwestern University.

Hayes, S. C., Brownstein, A. J., Haas, J. R., & Greenway, D. E. (1986). Instructions, multiple schedules, and extinction: Distinguishing rule-governed from schedule-controlled behavior. *Journal of the Experimental Analysis of Behavior, 46*, 137–147. doi: 10.1901/jeab.1986.46-137

Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon *Science, 146*, 549–551. doi: 10.1126/science.146.3643.549

Horne, P. J., Lowe, C. F., & Harris, F. D. A. (2007). Naming and categorization in young children: V. Manual sign training. *Journal of the Experimental Analysis of Behavior, 87*, 367–381. doi: 10.1901/jeab.2007.52-06

Horne, P. J., Lowe, C. F., & Randle, V. R. L. (2004). Naming and categorization in young children: II. Listener behavior training. *Journal of the Experimental Analysis of Behavior, 81*, 267–288. doi: 10.1901/jeab.2004.81-267

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

Keller, F. S., & Schoenfeld, W. N. (1950). *Principles of psychology*. New York: Appelton-Century-Crofts, Inc.

Laurence, S., & Margolis, E. (1999). Concepts and cognitive science. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 3–81). Cambridge, MA: MIT Press.

Lowe, C. F., Horne, P. J., Harris, F. D. A., & Randle, V. R. L. (2002). Naming and categorization in young children: Vocal tact training *Journal of the Experimental Analysis of Behavior, 78*, 527–549. Retrieved from http://www.ncbi.nlm.nih.gov.

Lowe, C. F., Horne, P. J., & Hughes, J. C. (2005). Naming and categorization in young children: III. Vocal tact training and transfer of function. *Journal of the Experimental Analysis of Behavior, 83*, 47–65. doi: 10.1901/jeab.2005.31-04

Margolis, E., & Laurence, S. (1999). *Concepts: Core readings*. Cambridge, MA: MIT Press.

Miguel, C. F., & Kobari-Wright, V. V. (2013). The effects of tact training on the emergence of categorization and listener behavior in children with autism. *Journal of Applied Behavior Analysis, 46*, 669–673. doi: 10.1002/jaba.62

Miguel, C. F., & Petursdottir, A. I. (2009). Naming and frames of coordination. In R. A. Rehfeldt & Y. Barnes-Holmes (Eds.), *Derived relational responding: Applications for learners with autism and other developmental disabilities*. Oakland, CA: New Harbinger.

Miguel, C. F., Petursdottir, A. I., Carr, J. E., & Michael, J. (2008). The role of naming in stimulus categorization by preschool children. *Journal of the Experimental Analysis of Behavior, 89*, 383–405. doi: 10.1901/jeab.2008-89-383

Palmer, D. C. (2002). Psychological essentialism: A review of E. Margolis and S. Laurence (Eds.), concepts: Core readings. *Journal of the Experimental Analysis of Behavior, 78*, 597–607. doi: 10.1901/jeab.2002.78-597

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of experimental Psychology: General, 104*, 192–233. doi: 10.1037//0096-3445.104.3-192

Rosch, E. (1999). Principles of categorization. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 189–205). Cambridge, MA: MIT Press.

Rosenfarb, I. S., Newland, C., Brannon, S. E., & Howey, D. S. (1992). Effects of self-generated rules on the development of schedule-controlled behavior. *journal of the Experimental Analysis of Behavior, 58*, 107–121. doi: 10.1901/jeab.1992.58-107

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

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

Sidman, M., Wynne, C. K., Maguire, R. W., & Barnes, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior, 52*, 261–274. doi: 10.1901/jeab.1989.52-261

Sigurdardottir, Z. G., Mackay, H. A., & Green, G. (2012). Stimulus equivalence, generalization, and contextual stimulus control in verbal classes. *The Analysis of Verbal Behavior*, *28*, 3–29. Retrieved from http://www.ncbi.nlm.nih.gov/.

Smeets, P. M., Dymond, S., & Barnes-Holmes, D. (2000). Instructions, stimulus equivalence, and stimulus sorting: Effects of sequential testing arrangements and a default option. *The Psychological Record, 50*, 339–354. Retrieved from http://opensiuc.lib.siu.edu/.

Smyth, M. M., Collins, A. F., Morris, P. E., & Levy, P. (1994). *Cognition in action* (2 ed.). UK: Psychology Press Ltd.

Vaughan, W. (1988). Formation of equivalence sets in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 36–42. doi: 10.1037//0097-7403.14.1.36

Zentall, T. R., Galizio, M., & Critchfield, T. S. (2002). Categorization, concept learning, and behavior analysis: an introduction. *Journal of the Experimental Analysis of Behavior, 78,* 237–248. doi: 10.1901/jeab.2002.78-237

The Use of Sorting as an Additional Measure on Stimulus Equivalence

Aina Dahle Norbom

Oslo and Akershus University College, Department of Behavioral Science

Sorting tests are seldom referred to in the research of stimulus equivalence. Tests for derived relations when studying stimulus equivalence are time consuming and require much effort from participants. In regards of different studies there could be of interest in many cases to use a sorting test as an additional measures on equivalence class formation. The sorting is conducted within a few minutes compared to the derived relations test that often takes a minimum of 30 minutes to finish. The test is also an easy measure for the experimenter in that methods for measuring the sorting results are developed for the use of computers in contrast to what has previously been done with paper-and-pencil or laminated cards. The present study examined in which degree a sorting test could serve as an additional measure on equivalence class formation. Participants were exposed to conditional discrimination procedures using a LS (linear series) training structure to train potentially three 5-member classes. Participants were tested for symmetry, transitivity, and equivalence in a MTS (matching-to-sample) test after training. Participants were presented with two phases of the experiment. Phase 1 was mandatory for all participants and then they were allocated to Phase 2 dependent on the results from Phase 1. The sorting tests were used as a pretest and a posttest during Phase 1 and as a pretest, posttest (after training) and post-posttest (after MTS-test) in Condition 1, Phase 2. Results showed that participants who responded in accordance with stimulus equivalence also sorted the stimuli in experimenter-defined classes. Participants who did not respond in accordance with stimulus equivalence did not sort the stimuli in experimenter-defined classes. Also, participants who had low scores on the MTS-test had less experimenter-defined classes in the sorting tests.

Keywords: sorting, categorization, equivalence class formation, matching-to-sample

The Use of Sorting as an Additional Measure on Stimulus Equivalence

In stimulus equivalence research sorting tests has been of interest in some studies. In other fields of psychology, e.g., cognitive psychology, categorization has played a central role for decades (Barsalou, 1992; Garner, 1978; Mervis & Rosch, 1981; Nosofsky, 1986, 1987; Rosch, 1973, 1999; Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Bream, 1976; Zentall, Galizio, & Critchfield, 2002). Barsalou (1992) suggested that categorization is a process of the determination of things that goes together. A class of stimuli or a group of stimuli can be said to be a category, and a concept can be said to be a categorization process that is facilitated by knowledge. Another definition of categorization is suggested by Rosch (1999); "to categorize a stimulus means to consider it, for purposes of that categorization, not equivalent to other stimuli in the same category but also different from stimuli not in that category" (pp. 190). In addition, Rosch (1999) states that knowing about properties and differentiating between them with fine discriminations would lead to categories in large numbers.

Stimulus equivalence and categorization can be closely related, but explanations and descriptions of what occurs in concept formation are slightly different in behavior analysis and cognitive psychology. In behavior analysis, experiments have been conducted within the field of stimulus equivalence focusing on sorting and equivalence class formation. Sorting occurs when participants sort stimuli into groups (Arntzen, Braaten, Lian, & Eilifsen, 2011; Eikeseth, Rosales-Ruiz, Duarte, & Baer, 1997; Eilifsen & Arntzen, 2009; Fields, Arntzen, Nartey, & Eilifsen, 2012; Green, 1990; Grimm, 2011; Pilgrim & Galizio, 1996; Smeets, Dymond, & Barnes-Holmes, 2000). Stimulus equivalence is defined in terms of three features; reflexivity, symmetry and transitivity e.g., Sidman (1994). Reflexivity means that A is related to A and B is related to B, Thus, the stimulus must be related to itself. Symmetry means that if A is related to B, then B is related to A. The stimuli change, thus, the stimuli are bidirectional. Transitivity means that if A is related to B, and B is related to C, then A must be related to C. These features define stimulus equivalence. Additionally, an important aspect of stimulus equivalence is that the stimuli have to be able to be replaced without interfering with the response. This means that stimuli is interchangeable (Green & Saunders, 1998). In regards of stimulus equivalence the three-term contingency is not sufficient and Sidman (1994) included a fourth control variable; a conditional stimulus controlling the three-term contingency. For example, if it is raining outside, the rain will be the conditional stimulus for the use of an umbrella. The umbrella will be the discriminative stimulus for using the umbrella, and *don't get wet* will serve as a reinforcer. If it were sunny outside the sun would serve as another conditional stimulus for a different three-term contingency.

When conducting studies on stimulus equivalence, the most commonly used procedures is conditional discrimination, and conditional relations is defined when participants respond to B1 in the presence of A1. These procedures can generate matching-to-sample (MTS) responding. Thus, before participants responding can be defined as MTS-responding the features of reflexivity, symmetry and transitivity have to be established through conditional discrimination (Sidman & Tailby, 1982). MTS-trials are mostly conducted with the presentation of a stimulus, either identical or arbitrary which is presented alone in the middle of a screen. Identity matching means that the sample stimulus and comparison stimulus are identical. When stimuli are arbitrary the sample stimulus and the comparison stimulus do not have any physical similarities, and Greek and Arabic letters are frequently used in experiments on stimulus equivalence. The stimulus that is presented in the middle of the screen is called a sample stimulus, and when the participant presses the sample stimulus, two or more new stimuli will appear on the screen. These stimuli are called comparison stimuli. One of the comparison stimuli is defined as correct by the experimenter, and the participant has to click on one of them in order to find out which one is correct. If participants chose the stimulus defined as correct, it will appear “good”, “fantastic” etc. on the screen. If the response is incorrect the feedback will state “wrong”. After each feedback, a new trial will follow until the mastery criterion for that particular experiment is achieved (Green & Saunders, 1998). Other procedures are applied in establishing stimulus equivalence in addition to conditional discrimination training. These contain respondent-like procedures or stimulus-stimulus pairing (Layng & Chase, 2001; Leader & Barnes-Holmes, 2001a, 2001b; Leader, Barnes-Holmes, & Smeets, 2000). Leader et al. (2000) examined the possibility of respondent-type procedures in establishing stimulus equivalence in young children. They used a linear series training structure along with one-to-many and many-to-one and testing for emergent relations were conducted using a standard MTS-procedure. The results showed that using respondent-type procedures across all three training structures produced equivalence responding and they concluded that this procedure certainly is applicable in establishing equivalence relations. However, Layng and Chase (2001) found it was not sufficient to use only stimulus-stimulus pairings in establishing equivalence classes, but the need for a contingency between the training, different alterations and instructions to the MTS-test for relations to be established.

Protocols is a variable effecting stimulus equivalence responding. Most widely used in stimulus equivalence research is the simultaneous protocol where all relations are presented from the start. In addition to the protocols, there are different training structures used in experiments e.g., linear series training structure (LS). In LS the relations are trained from A1 to B1, from B1 to C1, from C1 to D1 and so forth, dependent on the number of classes in the experiment (Green & Saunders, 1998). The linear series training structure is the most difficult training structure in the establishment of stimulus equivalence (Arntzen, Grondahl, & Eilifsen, 2010; Arntzen & Hansen, 2011; Green & Saunders, 1998). An emergent relations test is mostly used in equivalence research. Testing for stimulus equivalence with an emergent relations test can take from 30 minutes to several hours depending on the number of classes and members. Conducting a sorting test only takes minutes compared to the test for derived relations and is effortless and less time-consuming. Dymond and Rehfeldt (2001) proposed supplemental measurements in regards of derived stimulus relations and proposed evidence that a sorting test is a valid measurement of derived stimulus relations. Participants are provided with all the stimuli at once and can categorize or sort the stimuli into groups (Dymond & Rehfeldt, 2001). The sorting test can be applied in various different ways, e.g., paper-and-pencil (Smeets et al., 2000), with laminated cards (Fields, Arntzen, & Moksness, 2013; Fields et al., 2012) or as used in the present study on a computer using a computer mouse. The sorting test can be a more sensitive measure of equivalence class formation compared to the test for derived relations. More sensitive in this case means that participants have established experimenter-defined classes in the sorting test, but fails to responds in a accordance with stimulus equivalence in the test for derived relations. For example, participants can press the wrong stimuli in the test, which can result in responding not in accordance with stimulus equivalence. Additionally, participants can respond correctly during training, which indicates that conditional discriminations are established and get through to testing for derived relations. However, again they fail to respond in accordance with stimulus equivalence, but sort the stimuli in experimenter-defined classes. Thus, the sorting test is a more sensitive measure in the way that the classes are established in one certain test (sorting test) but not in another (standard test for derived relations).

There have been some studies that have used sorting tests as additional measures of equivalence class formation (Arntzen et al., 2011; Eilifsen & Arntzen, 2009; Fields et al., 2013; Fields et al., 2012). Fields et al. (2012) examined the effects of meaningful stimuli in the establishment of stimulus equivalence, and used a sorting test with 15 laminated cards as a supplemental measure of class formation. The sorting test was presented for the participants as a pretest (before training) and posttest (after derived relations test). The correct way to sort the stimuli would be to cluster five stimuli into three groups defined as correct by the experimenter. The results showed that the participants who responded in accordance with stimulus equivalence also sorted the stimuli in experimenter-defined classes in the posttest. A question that was asked in regards of the correlation between the two tests was if negative results on the test for derived relations would yield negative results on the sorting test. Three participants were examined for this purpose and the findings indicated that even though one participant sorted all the stimuli in class 1 correct, the participant chose comparisons from all three classes in the test for derived relations, and not only the comparisons that were part of class 1. This shows dissociation between the two tests and is an important observation (Fields et al., 2012). Arntzen et al. (2011) also used sorting as an additional measure of equivalence class formation. Participants were instructed to sort the stimuli before and after the standard test for stimulus equivalence. These results were deviant from the results of previous studies (Fields et al., 2012; Smeets et al., 2000) and partially the present study. The results showed that even though participants did not respond in accordance with stimulus equivalence in the derived relations test they sorted the stimuli in experimenter-defined classes.

In an experiment conducted by Eilifsen and Arntzen (2009), participants were instructed to categorize stimuli before and after a MTS-test. The study involved 20 participants and out of these, ten responded in accordance with stimulus equivalence. The study involved to parts; one test in which if participants responded correctly, the experiment was finished, and a second test for those who responded incorrectly on the first test. To qualify as responding in accordance with stimulus equivalence the participants had to respond correct in either of the two tests. 14 out of 20 participants categorized the stimuli in experimenter-defined classes. This yields that four out of the 14 participants that categorized the stimuli correct did not respond in accordance with stimulus equivalence on the derived relations test. This is partially in accordance with what was found in the present study, but not in accordance with Fields et al. (2012). Another study that examined sorting was Fields et al. (2013). The results showed 100 percent correspondence between the results on the test for derived relations and sorting test in 43 of 44 participants. Thus, when participants responded in accordance with stimulus equivalence on the derived relations test, they sorted the stimuli in experimenter-defined classes and vice versa. Additionally, participants who did not respond in accordance with stimulus equivalence but had one established class in the test for derived relations sorted these stimuli together in the sorting test.

Smeets et al. (2000) conducted a study with a sorting test in addition to the derived relations test. Experiments 2 and 3 included a sorting test after the derived relations test to measure if the results from the two tests correlated. The experiments consisted of paper-and-pencil formats when conducting the tasks. The instruction and examples was placed on page 1, page 2 contained rules on how the tasks worked, and the third page consisted of 32 different tasks. After memorizing the rules from the instruction, they had to hand in the page to receive page 3. Here there was some of the same tasks as from page 2 and some tasks which were new. On page 4 in Experiment 2 a new test, similar to the one on page 3, were presented and participants were instructed to complete all trials. Page 5 instructed participants to categorize the stimuli into two groups. The stimuli were placed on the left side in a column and in the right column “group 1” or “group 2” was printed, and participants had to draw lines. In Experiment 3 in Smeets et al. (2000) the method was similar to Experiment 2, except that in Experiment 3 the instruction stated that participants could skip difficult tasks (pages 3 through 5). The overall results showed that number of derived relations and the results from the sorting test covaried. So, when participants had more correct responding on the derived relations tests, more correct sorting was shown. They concluded the sorting test in regards of this particular experiment not to “provide convergent validity for stimulus equivalence”

(Smeets et al., 2000, p. 352). They also stated “a sorting test could be seen as a functional equivalence test without training but with an instruction to classify, and in which the order in which the stimuli are grouped is determined by the subject” (p. 352). Further they stated that knowledge on how functional equivalence and sorting covaries may give rise to new classification procedures in regards of derived relational responding.

The present experiment was conducted in a slightly different manner concerning the sorting test, which was conducted on a computer in contrast to the Smeets et al. (2000) study using paper-and-pencil formats and in contrast to Arntzen et al. (2011), Fields et al. (2013) and Fields et al. (2012) using laminated cards. The procedure in the current study gave the participants easy access to the stimuli and they could be moved around the screen effortlessly and be easily changed to form groups. Most of all it gave an easy access to measure what participants sorted by taking a screenshot of the completed sorting. Further, reaction time is a variable that has been studied in several other studies (Arntzen, 2004; Arntzen et al., 2011; Arntzen & Lian, 2010; Bentall, Dickins, & Fox, 1993; Eilifsen & Arntzen, 2009; Holth & Arntzen, 1998; Spencer & Chase, 1996). Measuring reaction time can give an indication of the strength of a relation and that the descriptions of that relations are more precise (Spencer & Chase, 1996). In addition it can give an indication of a correlation between reaction time and the different trial types and reaction time and the number of correct responses on the test for stimulus equivalence. Previous studies have shown results indicating lower reaction time in the first trials during test, thus, the reaction time decreases during the last trials. (Arntzen & Hansen, 2011; Eilifsen & Arntzen, 2009; Spencer & Chase, 1996; Wulfert & Hayes, 1988). These studies have also shown higher reaction time during symmetry trials compared to the directly trained trials, and those trials of transitivity and equivalence yields higher reaction time than symmetry trials.

The present study sought to study in which degree there was any correlation between derived relations responding and sorting responding. In that case a sorting test would prove more applicable as an additional measure of equivalence class formation. To clarify, the correlations are not calculated between the two different tests, thus, the term will serve as an indication of congruence between a sorting test and an emergent relations test. In addition the present study examined reaction time and the correlation between the reaction time and the types of trials, and reaction time and number of correct responses.

Method

Participants

The participants were students and persons with different occupations aged from 20 to 43 years old, averaging 28.4, whereas seven were females and nine were males. The students attended a college health degree and the other participants were requited through personal contacts. None of the participants had been part of a similar experiment before and had no pretraining or knowledge of stimulus equivalence.

Setting and apparatus

The experiments were conducted in an office approximately 3 x 4 meters. One third was arranged as a lab cubicle where participants sat during the experiment. There were no distracting elements in the cubicle. The cubicle comprised one desk with a computer, and a chair for the participants, which faced a blank white wall. While the experiments were conducted the experimenter sat outside the office to avoid questions about the tasks, and interrupting the participants.

The computer was a HP EliteBook 8760w Intel Core i5-2540M (2,6GHz). The same computer was used during all phases and conditions including the sorting tests. The program used for the matching-to-sample (MTS) tests was Cognitive Science Partner and the sorting program was developed specifically for the experiment by a member of the lab. The computer screen was 36.8 cm in width, 23 cm in height and 43 cm in diagonal. It measured 16 cm from the left of the screen and in to the stimulus and 15.8 cm from the right and in to the stimulus. From the top and from the bottom and in to the stimuli it measured 9 cm. The stimuli were 5 cm high and 5 cm wide and solid black. The measures were taken from the outline of the stimuli. Thus, as far out on as the screen reacted when pushing the mouse button on the lines of the stimulus. See Figure 1 for the stimuli used in Phase 1 and Phase 2, Condition 2 through 4, and Figure 2 for the stimuli used in Phase 2, Condition 1.

Design

A single subject research design with different conditions was used. The sorting test worked partially as a baseline measure. Thus, it worked as a baseline for equivalence class formation on the consecutive sorting tests where the training in the MTS-program was the intervention. The control in the experiment was the reinstatement of conditions contingent on responding, thus, the conditions had different manipulations in regards of responding.

Procedure

General information to participants. Upon arrival all participants received a written consent form, which they were told to read and sign. Participants were informed briefly about stimulus equivalence and the experiment they were about to participate in through the consent form. They were informed when recruited that the experiment could take up to four hours. When finished reading and signing the consent form, participants were asked if they had any questions, and were told that they would find out what to do through the course of the experiment. Additionally, they were told they were anonymous and withdrawal from the experiment was possible at any time.

Dependent variable. The dependent variable was the number of correct responses on the MTS-test and correct or incorrect on the sorting test. Correct responding on the sorting test was defined when participants sorted stimuli in experimenter-defined classes.

Additionally, reaction time was measured. Defining properties of reaction time was duration from when the sample stimulus was pressed until the comparison stimulus was pressed. Instead of using ordinary reaction time, the measure was inversed reaction time. Spencer and Chase (1996) defined this as inversed latency, but the term inversed reaction time will be applied in the present article.

Conditions. The experiments consisted of Phase 1 and Phase 2 with together five conditions. In Phase 1 all participants underwent the same condition. Phase 1 consisted of a sorting test (pretest), followed by training in the MTS-program and a following test for symmetry, transitivity and equivalence (derived relations) in the MTS-program, ending with an additional sorting test (posttest). Contingent on the results from the derived relations test and the sorting test in Phase 1, participants were allocated to one of four conditions in Phase 2. The allocation depended on the test results from Phase 1. For an overview of the conditions see Figure 3. Following will be a detailed description of the conditions.

Condition 1. If participants responded in accordance with stimulus equivalence on the test for derived relations and the sorting test in Phase 1 they were allocated to Condition 1. This condition consisted of a sorting test (pretest), training in the MTS-program followed by a new sorting test (posttest), new test for derived relations in the MTS-program, and a last sorting test (postposttest). If participants did not sort in experimenter-defined classes in the last sorting test and neither responded in accordance with stimulus equivalence, a new test for emergent relations, and a new sorting test was presented. If participants did not form equivalence classes on the sorting test, but responded correctly on the derived relations test, then only a new sorting test was presented.

Condition 2. If participants did not responded in accordance with stimulus equivalence on the test for derived relations but sorted the stimuli in experimenter-defined classes, a new sorting test and test for emergent relations was repeated with the same stimulus set.

Condition 3. If participants responded in accordance with stimulus equivalence on the test for derived relations, but did not sort the stimuli in experimenter-defined classes, then a sorting test was repeated with the same stimulus set.

Condition 4. If participants did not respond in accordance with stimulus equivalence on the test for derived relations or sorted the stimuli in experimenter-defined classes, a new derived relations test and sorting test was reintroduced with the same stimulus set.

Sorting test. When participants signed the consent form, they were presented with the sorting test presented on the computer. Participants were given the instruction "put these into groups as you feel like" in the presence of the stimuli presented on the screen. The stimuli were placed as a deck of cards. Participants were informed that in order to see the next stimulus the one top had to be moved. The stimuli were black and presented on a white computer screen and they could be moved around using a computer mouse. When the participant was finished, he or she called the experimenter and a screenshot of the sorting results was taken.

Matching-to-sample. After the sorting, participants were presented with the MTS-program. They were asked to push the start button when ready, and after pushing "start" an instruction for the MTS-training was shown in Norwegian, stating the following translated into English:

> A stimulus will appear in the middle of the screen. Click on this by using the computer mouse. Three other stimuli will then appear. Choose one of these by using the computer mouse. If you choose the stimuli 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!

The training protocol was simultaneous, and the training structure was linear series. Participants were trained from A->B->C->D->E. Thus, if A1 is presented as a sample stimulus, B1 is the correct comparison. See Table 1 for an overview of how the relations were trained and the number of trials per relation in training and test. The stimuli was presented concurrent, thus, all the relations were presented mixed in each block. The experiment used 15 different stimuli, three classes with five members each. This gave 60 different relations to be learned; AB, BC, CD and DE. In the training trials, each block had 60 relations. First, the participants went through training with feedback on each trial. The feedback stated “great”, “good”, “or excellent” etc when responses were correct or “wrong” whenever responses were incorrect as defined by the experimenter. To pass through training, correct responding had to exceed 90 percent correct of the trials in each block. If not, the blocks were enrolled. When responding within the criteria on one block, a new block was presented. The new block consisted of feedback fading from 100 to 75 percent of the trials. When that block was on 90 percent correct or more a block with 25 percent feedback were presented, and then zero percent feedback. With 90 percent correct or more on the zero percent feedback blocks the test for derived relations was presented.

In the MTS-test participants were tested for symmetry (BA, CB, DC and ED), transitivity (AC, AD, AE, BD, BE and CE) and equivalence (EA, EB, EC, DA, DB, CA). The test consisted of 300 trials, divided into 60 baseline trials, 60 symmetry trials, 90 transitivity trials and 90 equivalence trials. To respond within 90 percent correct of the trials in test, participants had to respond correct on 54 out of 60 on the baseline and symmetry trials and 81 out of 90 on the transitivity and equivalence trials. When finished with the last sorting test, participants had a debriefing with the experimenter where they were presented with their result files from the MTS-program and the sorting program.

Results

The results show 16 participants out of 17 who participated in the study. One participant was excluded from the experiment due to alterations in the procedure after the participant had conducted the experiment. This participant was exposed to a different procedure containing a sorting test in between training and test in the MTS-program.

Three participants, P5203, P5207, and P5210 responded correct in Phase 1 on the test for derived relations and the sorting test, and all three were allocated to Condition 1 in Phase 2. The participants also responded correct on the test for derived relations and on the two posttests of sorting in Phase 2. When comparing the results from the two testes in both phases of the experiment it almost shows a correlation of 100 percent. P5203 and P5207 did not respond correct on all relations in the MTS-test, a few random errors was measured. This was the same with P5210 in the test for equivalence in Phase 1. However, they formed experimenter-defined classes in the sorting tests.

Out of the 13 participants who did not respond in accordance with stimulus equivalence in Phase 1 only three sorted the stimuli in experimenter-defined classes (P5202, P5204, and P5212). P5202 and P5212 sorted in experimenter-defined classes in Phase 1and 2 and responded in accordance with stimulus equivalence in the emergent relations test in Phase 2. Both P5202 and P5212 had a 100 percent correlation between the test for derived relations and the sorting test in Phase 2. They also showed a higher correlation between the symmetry trials and the sorting results in Phase 1. P5204 sorted in experimenter-defined classes in Phase 2 but did not respond in accordance with stimulus equivalence in either of the phases. He had a higher correlation between the symmetry trials and the sorting results in Phase 1 and Phase 2. The remaining 10 participants did not respond in accordance with stimulus equivalence in the MTS-program nor sorted the stimuli in experimenter-defined classes in either of the phases.

In the emergent relations test in Phase 1, seven participants had the directly trained (DT) relations intact (P5202, P5204, P5208, P5211, P5212, P5213 and P5214). Out of these P5202, P5208, P5211 and P5212 also had the symmetry trials intact during. In Phase 2 (Condition 2) only P5202, P5204, P5211 and P5212 still had the DT relations intact (P5202 and P5212 responded in accordance with stimulus equivalence). Two participants only had the symmetry (SY) relations intact during the tests, P5201 in Phase 1 and P5208 in Phase 2. Table 2 shows a detailed overview of participant results from the emergent relations test and sorting tests and whether participants sorted in experimenter-defined classes. A result common for each participant was the correlation between the way they sorted the stimuli and which trial types that was intact in the derived relations test. For example if participants had only baseline and symmetry trials intact during the derived relations test, then these relations were sorted together in the sorting test. Similarly, participants who formed equivalence classes in the sorting test and responded in accordance with stimulus equivalence had almost a full correlation between the two tests. The participants who responded with low scores in the tests for derived relations showed a lower correlation between the sorting test and the emergent relations test.

P5201, P5204, P5205, P5206, P5209, P5211, P5213, P5214 and P5216 showed low scores on the derived relations tests, especially on the tests for transitivity and equivalence. Only P5211 responded over 50 % (45/90 correct or above) correct on the equivalence trials on the test in Phase 1 and Phase 2, the remaining eight participants responded below 50 % correct on the test trials for transitivity and equivalence. Common for these participants was low scores on the emergent relations test and also the correlation between the sorting test and the emergent relation responding was low. P5202, P5208, and P5212 showed slightly different results. They had higher scores on transitivity and equivalence trials and also showed a higher correlation between sorting and the results from the derived relations test on both symmetry and equivalence. Additionally, baseline trials and symmetry trials were intact during the tests in both phases (except P5208 who did not have baseline trials intact during the test in Phase 2) and the highest correlation was between the sorting test and responding on symmetry trials.

Figure 4 and 5 shows an example of the correlation between the sorting and the results on the symmetry trials and equivalence trials for P5211 and P5215. The matrix shows which comparison participants choose in the presence of a certain sample stimulus. The stimuli in the most left column are the sample stimuli B through E. The stimuli in the top row show comparison stimuli A through D. For participant to get all relations correct, he would need to have all boxes filled with the number “5” in the diagonal. As in this example of P5211, she had a few outsiders in the test for symmetry e.g. she chose comparison A2 two times in the presence of sample B3, whereas A3 was the correct comparison. In regards of sorting, the different grey tones in the boxes represent participant sorting. The darkest filled boxes are the trials the participant responded correct on and also sorted together in the sorting test. The lighter grey color with the number “5” indicated that the participant responded in accordance with symmetry, but not sorted the stimuli together. The boxes with the lightest shade of grey were the ones where the participants sorted the stimuli together but did not respond correct on all five relations. The boxes without filling indicate incorrect trials and those stimuli were not sorted together in the sorting test. To summarize, the more dark boxes and lightest shade of grey indicate a greater correlation between the sorting test and the outcomes on the emergent relations test.

Figure 4 shows the tests for symmetry and equivalence in Phase 2 for participant 5211. The top panel shows a matrix of symmetry trials and the bottom panel shows a matrix of equivalence trials. P5211 responded in accordance with symmetry trials in both tests for derived relations and additionally sorted these stimuli together in the sorting test. As shown in the top panel of Figure 4 the greater number of darkest boxes indicates a high correlation between the sorting test and derived relations test. The lower panel of Figure 4 shows a low correlation between the sorting test and equivalence trials from the test for derived relations since many boxes are unfilled and the numbers are outside of the boxes in the diagonal.

P5215 was the only participant with results different from the other participants. He did not respond in accordance with any of the test trials in either of the two phases and in addition the scores were low on every test trial in both phases. However, he still had a high correlation between the sorting test and tests for both symmetry and equivalence in Phase 2. Figure 5 shows matrix showing the correlation between the sorting test and symmetry test (upper panel) and between the sorting test and equivalence test (bottom panel). The matrix shows almost 100 percent correlation between sorting and symmetry trials and 100 percent correlation between the sorting and the equivalence trials, even though he's scores on the emergent relations test shows he did not respond in accordance with any of the test trials.

None of the 16 participants sorted the stimuli in accordance with experimenter-defined classes on the first sorting test conducted in Phase 1 (pretest). Table 3, 4 and 5 shows the number of participant-defined classes established in the presorting, postsorting and postpostsorting test respectively. It shows the number of classes participants formed during sorting tests. However, it fails to demonstrate which stimuli participants sorted into each group.

Figure 6 and 7 display results of reaction time for all participants. The figures show the graphs of the five last trials during training, which is marked "training", the five first and five last trials through all test relations (baseline, symmetry, transitivity and equivalence). Additionally, it's divided into Phase 1 and 2. The graphs also show correct and incorrect responding. The solid black bars indicate correct responses while the grey bars indicate incorrect responses. Thus, of the five relations for each trial type, bars could either be all black, all grey, or split dependent on how participants responded. The reaction time is reversed reaction time and the formula is as follows: 1/RT equals inversed reaction time. When analyzing the reaction time it will be reversed from the original, so when different relations have high bars that illustrates short reaction times and vice versa.

Discussion

The purpose of the current study was to examine if there was a correlation between results from tests for derived relations and results on sorting tests, and perhaps increase the sorting applicability as an additional measure on equivalence class formation. A point that needs attention is the difference between the sorting test and the test for derived relations. It is not possible until further examination has been conducted, to conclude that these two tests measure the same thing. The two tests as a starting point measures not the same variables in that they are conducted differentially. Supplementary, the current study sought to find differences or correlations between reaction time and number of correct responses on the derived relations test. Thus, whether there could be coherence between longer reaction time and non-accordance with stimulus equivalence and vice versa. Results showing faster reaction time when responding is correct indicate the strength of the response.

Test Correlation

Overall results showed a correlation between responding on the test for stimulus equivalence and sorting test. It was higher between the sorting test and the test for emergent relation when participants responded more correct on the different relations. Thus, when participants responded in accordance with stimulus equivalence on the emergent relations test and in addition yielded high scores, participants sorted the stimuli in experimenter-defined classes. In regards of participant 5203, 5207, and 5210, all of them consequently sorted the stimuli in experimenter-defined classes and responded correct on more than 90 percent of the test trials in the derived relations test. Except for three incorrect responses in the first test and four in the second test for derived relations, participant 5207 sorted the stimuli in experimenter-defined classes. P5203 and P5210 also displayed incorrect responses, and these were most likely random errors. Comparing sorting results, which shows that participants established equivalence classes, argumentation is likely that the errors on the test for derived relations are random errors. The results from these three participants in particular makes sorting tests more applicable as an additional measure on equivalence class formation in regards of the absolute conformity between the results from the two tests. A positive side to sorting tests is the chance of making errors is low when participants have established the experimenter-defined classes. As predicted, no participant responded in accordance with stimulus equivalence but at the same time failed to sort the stimuli in experimenter-defined classes.

Stimulus Equivalence Responding vs. Equivalence Class Formation Responding

Interesting and prospective results was if participants did not respond in accordance with stimulus equivalence on the test for derived relations, but sorted the stimuli in experimenter-defined classes. P5202, P5204 and P5212 responded in this manner. P5202 and P5212 sorted the stimuli in experimenter-defined classes after the test for derived relations in Phase 1, while participant 5204 did so after the test in Phase 2. Interestingly, P5204 did not respond in accordance with stimulus equivalence on either of the tests for derived relations. Results from P5202, P5204 and P5212 indicate that a sorting test is a more sensitive measure than the test for stimulus equivalence. These results are consistent with the results found in other studies (Fields et al., 2013; Smeets et al., 2000). In congruence with the results found in Fields et al. (2013), when most participants in the present study failed to respond in accordance with stimulus equivalence the sorting was not in accordance with experimenter-defined classes. The current study found results where participants sorted 100 percent correct but did not respond in accordance with stimulus equivalence on the MTS-test. In contrast, Fields et al. (2013) found a correlation between the classes established in the MTS-test (e.g. one class) and the sorting results. They did not find participants that sorted the stimuli 100 percent correct when responding was not in accordance with equivalence in the emergent relations test, which was found in the present study. The current study used a sorting test prior to the standard test (Phase 2, Condition 1) as suggested by Fields et al. (2013). P5203, P5207, and P5210 sorted the stimuli correct on the sorting test after training and showed an almost full correlation between the two different tests for class formation. It would be interesting to measure if participants would show the same results in the procedure in Condition 2 as in the procedure of Condition 1. Thus, it would be of interest if sorting responses was correct after training, but not stimulus equivalence responding, when presenting the sorting test first.

Controlling Variables and Correlation between Tests

The participants that failed to respond in accordance with stimulus equivalence and additionally yielded low scores in the test for derived relations did not display high correlation between the emergent relations test and sorting test. This indicates another variable controlling the participant’s response. As with P5215 shown in Figure 5, the results showed a 100 percent correlation between tests even though he did not respond in accordance with stimulus equivalence or sorted the stimuli in experimenter-defined classes. Sidman and Tailby (1982) argued that participants do not display MTS-responding until they respond in accordance with the three features of stimulus equivalence; reflexivity, symmetry and transitivity. On the background of that argument P5215 is not displaying MTS-responding. However, one must point out that P5215 sorted in three groups (though not experimenter-defined) and also choose the same stimuli (as in the sorting) in the presence of one another in the test for derived relations. Argumentation that P5215 is not displaying MTS-responding can be argued, but the consequent selection and 100 percent correlation between tests must be pointed out. In Phase 1, the data from the test for derived relations and sorting test showed a small correlation and the stimuli were sorted into four groups in the sorting test. However, in Phase 2 on the derived relations test, as shown in Figure 5, he responded correct on seven different relations (C1A1/D1A1/E1C1/E1A1, E2A2 and D3B3/E3A3) and incorrect on 11 relations. Additionally, the incorrect responding on the 11 different relations was anyhow chosen consequently by P5215 (the number "5" in each box). An interesting point is that he got no feedback and still showed such a big difference in both the sorting test and the test for derived relations from Phase 1 to Phase 2. The control of the participant's behavior lies not in the conditions of the conditional discrimination procedure defined by the experimenter. The control lies somewhere else which one can only speculate. McIlvane and Dube (1992) proposed the terms "stimulus control shaping" and "stimulus control topography" to be reintroduced in behavior analytic literature. "Stimulus control shaping" refers to changes in the control of stimulus-response relations and was originally a term used for its shaping of stimulus control. "Stimulus control topography" refers to functional *stimulus* class in contrast to functional *response* class and it refers to the differentiating between members in that stimulus class. As with P5215 in the present experiment, the results can be seen in the light of this definition of procedures. Since the procedures of sorting tests, training in the MTS-program and the test for derived relations are different, they all serve different controls over behavior. The sorting procedure serves one type of control and it is the first procedure presented in the present experiment. The training serves another control and the test for derived relations serves a third. As mentioned before the control may not shift between procedures but rather lie in the procedure presented first, and this could have been the case with P5215. Even though he responded within the criteria of training, he still had low scores on the test for derived relations, but still sorted the relations together as responded in the test for emergent relations. Argumentation can be made that the control of his behavior did not shift from the procedures of the sorting test to the MTS-training. Thus, the control over his responding throughout all the procedures of the experiment was from the instructions or experimental procedure of the sorting test. Overall, the results of P5215 show that it is possible to find correlations between two different tests even though the responding is not in accordance with the criteria of either of the two types of tests. In that regard, stimulus sorting is even more applicable.

For all participants that failed to respond in accordance with stimulus equivalence and fail to sort the stimuli in experimenter-defined classes, it could be argued that other variables was controlling their responses as with P5215. An interesting part was that participants with very low scores on the derived relations test showed less correlation between the sorting test and the test for emergent relations. For example, in Phase 1, participant 5214 sorted the stimuli into seven classes with two members in each group except one group with three members. In the second sorting test the participant sorted stimuli into two groups. Even if there were only two groups, the correlation between the sorting and the standard test was still low. The participant had even four relations that he chose consequently throughout the emergent relations test, but did not sort together. As mentioned above in regards of P5215 this could have something to do with the variables controlling the behavior. For example the instructions can exert much control over responses compared to the conditional discrimination procedure. The instructions could have led participants to think that the sorting test was one test and the emergent relations test was another and that they directly did not have anything to do with each other. The instruction for the sorting test was given oral prior to the presentation of the sorting test, and the instructions for the MTS-program was written and a part of the MTS-program. For future research the instructions could be presented written on the screen and linked to the sorting program, which could give the test more fluency and also be a part of the general instruction that were presented in the written consent form. P5209 used longer time on the sorting task because he thought it would present feedback, which was described in the written consent form, but not in regards of the sorting test. Sidman (1994) proposed that instructions could have effect on the results of the derived relations test. Saying that stimuli “belong together” or “match” could alone make the participant respond in a certain way and not because of the experimental procedure itself. This could be the case for participant 5209 in which the instructions for the sorting test “put these into groups as you feel like” and that he was told in the written consent form that there would be presented feedback on right and wrong responses. Participant 5209 had practically the same sorting results through every sorting task, even presorting, although he got through the training and testing in the MTS-program. This could indicate that the participant followed instructions to the sorting test whenever he was instructed to sort the stimuli and did it on the basis of the first sorting he conducted in the pretest.

Dismissal of Order Effects between Sorting Tests and Derived Relations Tests

For participants in Condition 1 in Phase 2, they were presented with a different condition then in Phase 1. Thus, a sorting test in between training and the derived relations test were presented. It can be argued that sorting tests can have potential order effects on the responding in the MTS-test (Cooper, Heron, & Heward, 2007), but still these participants already responded in accordance with stimulus equivalence on all tests in Phase 1. Thus, participants were acknowledged with the procedure from Phase 1 and relations was established fast which indicates that the sorting test does not facilitate any responding in the MTS-test. On the contrary, P5203, P5207, and P5210 sorted in experimenter-defined classes before the MTS-test and indicate that sorting tests can be applied as a measure for equivalence class formation. Results like these also indicate that the sorting results are not facilitated by MTS-responding. This means that the results of the sorting test are not a product of the results on the test for derived relations. The participants do not need to undergo an emergent relations test to sort the stimuli in accordance with equivalence class formation. These types of data have not been found in the previous studies regarding sorting tests in addition to the standard test within the field of behavior analysis and stimulus equivalence.

Sensitivity of Sorting Tests

When analyzing the results they can indicate that a sorting test is a more sensitive measure of equivalence class formation as mentioned above. Fields et al. (2012) also found results similar to the present study. When participants formed equivalence classes in the emergent relations test, the sorting test also showed class formation in accordance with experimenter-defined classes. Fields et al. (2012) further argues that predictive results if the sorting test was presented before the emergent relations test, would probably show the same results. If that is the case it makes the sorting test more applicable as an additional measure on equivalence class formation. Sorting results can be compared with results from derived relations tests to measure if there is higher correlation between the results from the tests for symmetry or equivalence and the sorting test. In that case the results from the derived relations test are needed in addition. This was done and is shown for P5211 and P5215 in Figure 4 and 5. These results indicate that when participants have baseline and symmetry trials intact, they sort the stimuli in a way that conforms to the responding on the symmetry trials, but deviates from the equivalence trials. Thus, there was ha higher correlation between the results on symmetry trials and the sorting test, than on equivalence trials and the sorting test when the participants had symmetry trials intact during the test for derived relations but not the equivalence trials. This supports previous findings that a sorting test could be an additional measure of stimulus equivalence (Fields et al., 2013; Fields et al., 2012; Smeets et al., 2000).

The results of P5202 and P5212 especially shows that sorting could be a more sensitive measure, and a reason is the probability of making random errors in the test for derived relations is higher than in the sorting test. In the sorting test the participants can overview their sorting response before telling the experimenter they're finished. In the test for derived relations errors cannot be adjusted which they can in the sorting test. Thus, P5202 and P5212 had few errors, but still ha to conduct the test once more with the same stimulus set and not go through to Condition 1 in Phase 2. If the sorting test had determined the outcome, then they would have been allocated to Condition 1 in Phase 2 and not Condition 2 in Phase 2. Prospective procedures like these can be subject to further research using only sorting tests to measure equivalence class formation compared to the use of only MTS-tests.

To further argue that sorting could be a more sensitive measure, result files were examined to check for more correct responses at the end of the tests for derived relations, which could indicate delayed emergence (Sidman, 1994). This was not the case. There were not found any clear consistency between correct and incorrect responses through the first and second part of the experiment. The correct and incorrect responses were spread randomly through the test and showed no clear indication of more correct responses at the end. It was not nor any indication of delayed emergence in the sense that participants got higher scores on the test for derived relations in Phase 2. This is in contrast to Sidman's (1994) proposal stating that the test for derived relations could be necessary for the classes to be established.

The Role of Linear Series Training Structure

In Fields et al. (2013) almost half of the participants did not respond in accordance with stimulus equivalence. The majority of participants from the current study did not respond in accordance with stimulus equivalence. In both studies a linear series training structure was used, and can be a reason for why many failed to respond in accordance with stimulus equivalence and not sort the stimuli in experimenter-defined classes. This because studies have shown that this structure is the hardest when establishing equivalence classes (Arntzen et al., 2010; Arntzen & Hansen, 2011). The purpose of using this structure in the present experiment was because a sorting test will show more sensitive if the results in the test for derived relations is negative and sorting test results is positive.

Analysis of Reaction Time

When analyzing the reaction time of the participants, most of them had higher reaction time on incorrect relations and lower reaction time on correct relations. This indicates that reaction time says something about the strength of the relations. So, the lower reaction time the greater the strength. In regards of P5215 this was the case only that he had a higher reaction time on the correct responses than on the incorrect ones. This indicates the same since he defined he’s own classes and chose them together consequently. Thus, the strength of the relations matched his responding even though the reaction time was lower on the incorrect responses. These results have positive gain in the argumentation of the sorting test as an additional measure because the strength of the relations the participant defined was great and there was a full correlation. This shows that it is possible to use the sorting test as a measure in different test situations. For example if future experiment needs to be conducted on a shorter time span, sorting tests could be a measure alone for certain experiments. Off course all dependent on the purpose of the study.

The results of the inversed reaction time was mostly in accordance with the results from Spencer and Chase (1996). The current study also found that reaction time on baseline trials and symmetry trials were faster than on transitivity and equivalence trials. For each participant the results were almost the same. The inversed reaction time on the baseline trials and the symmetry trials showed the same pattern across participants. Either the baseline trials or the symmetry trials showed highest reaction time. There was no systematic pattern in which condition that yielded highest reaction time on baseline trials or symmetry trials, they varied unconditionally. These differences could be due to in which degree the relations where established by each individual participant. The same pattern was found for transitivity relations and equivalence relations. These were also variable throughout the experiment for each participant. On some occasions transitivity trials yielded highest reaction time, and on other occasions equivalence trials yielded highest reaction time without any systematic differences. Some of these variations can be due to other variables that may have disturbed the participant e.g. personal activity, drinks etc. Figures 6 and 7 show both incorrect and correct responses and another pattern visible was that for most participants, incorrect responses yielded lower reaction time or in some cases almost the same. These results are similar to several other studies regarding reaction time (Arntzen et al., 2011; Arntzen & Hansen, 2011; Sidman, 1994). Thus, when participants showed low reaction time on relations that were incorrect, the definitions of a functional relation was not present. So, if participants respond to a certain relation with low reaction time, the experimenter can predict the response as incorrect. The strength of the relation was low when the reaction time was high and the response was incorrect.

Sorting as a Measure of Equivalence Class Formation

The current study expands the knowledge about sorting tests in how it was conducted in a different manner than previous studies concerning sorting tests as an additional measure for equivalence class formation (Arntzen et al., 2011; Eikeseth et al., 1997; Grimm, 2011; Smeets et al., 2000). These previous studies used either paper and pencil (Eikeseth et al., 1997; Smeets et al., 2000) or laminated cards (Arntzen et al., 2011; Fields et al., 2013; Grimm, 2011). Even though these procedures had many benefits the present study showed additional advantages. The sorting tests were easier to conduct in regards of not having to use extra effort and time to make laminated cards, easier to take a screenshot of the finished sorting and have easy access to the results after experimentation. Additionally, it was extremely easy for participants to move stimuli around the screen as they wished. Easy possibilities for moving the stimuli around made it easy for participants to undo previous sorting and make new categorizations of the stimuli if wanted. Last but not least, the sorting program was an advantage when it came to time effort. This program made it effortless and time saving for the participants as well as for the experimenter. The stimuli were placed like a deck of cards because the experimenter did not want to give participants any indication about the number of classes and members that was defined as correct before starting the experiment. If participants had seen all the stimuli at once already placed in three rows and five columns or vice versa, that could have indicated the stimuli should be sorted into three 5-member classes or opposite. Whether this could have effect on the sorting results should be tested in future studies by comparing groups with different stimulus presentations. The most common way to present stimuli is in a pile and not spread out.

Practical implications of comparing sorting tests with the derived relations test is present since these two tests do not measure exactly the same variables. It needs further examination with comparing of results from both derived relations test and sorting test. If results continue to show conformity between how participants respond in the sorting test and in the test for derived relations that will decrease practical implications considerable. The use of sorting tests depends on the purpose of what is being studied. A derived relations test cannot be dismissed when the purpose of the study is to examine stimulus equivalence responding.

Limitations

A possible limitation with the present experiment was that it did not state anything about the sorting test in the written consent form, which then may have come as a surprise to participants who did not link the standard test and the sorting test together. Future research should combine these two and possibly make alterations with the instructions. A possible alteration should be to inform participants in the same consent form about the two different tests. Thus, not have two separate instructions and not one oral and one written. One common instruction should be compared with the procedures of the current study to examine whether it could have any effect on the results.

Conclusion

The present study mainly focused on the sorting test as an additional measure of stimulus equivalence and found that this type of test may indeed be used as a supplemental measure. The results showed that this test in most cases correlate with the results on the test for derived relations when participants respond in accordance with stimulus equivalence. Future research should definitely study this topic more thorough so that it could be used in the research on equivalence class formation.

References

Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *The Psychological Record, 54*, 275–291. Retrieved from http://opensiuc.lib.siu.edu/.

Arntzen, E., Braaten, L. F., Lian, T., & Eilifsen, C. (2011). Response-to-sample requirements in conditional discrimination procedures. *European Journal of Behavior Analysis, 12,* 505–522. Retrieved from https://oda.hio.no.

Arntzen, E., Grondahl, T., & Eilifsen, C. (2010). The effects of different training structures in the establishment of conditional discriminations and subsequent performance on tests for stimulus equivalence. *The Psychological Record, 60*, 437–462. Retrieved from http://opensiuc.lib.siu.edu/.

Arntzen, E., & Hansen, S. (2011). Training structures and the formation of equivalence classes. *European Journal of Behavior Analysis, 12*, 483–503. Retrieved from https://oda.hio.no.

Arntzen, E., & Lian, T. (2010). Trained and derived relations with pictures as node. *The Psychological Record, 60*, 659–677. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Barsalou, L. W. (1992). *Cognitive psychology: An overview for cognitive scientists*. Hillsdale, New Jersey: Lawrence Erlbaum Associates

Bentall, R. P., Dickins, D. W., & Fox, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *Quarterly Journal of Experimental Psychology, 46B,* 187–214.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied Behavior Analysis* (2 ed.). Upper Saddle River, New Jersey: Pearson Education, Inc.

Dymond, S., & Rehfeldt, A. (2001). Supplemental measures of derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin*, *19*, 8–12. Retrieved from http://psy.swan.ac.uk/.

Eikeseth, S., Rosales-Ruiz, J., Duarte, A., & Baer, D. M. (1997). The quick development of equivalence classes in a paper-and-pencil format trough written instruction. *The Psychological Record, 47*, 275–284. Retrieved from http://opensiuc.lib.siu.edu/.

Eilifsen, C., & Arntzen, E. (2009). On the role of trial types in tests for stimulus equivalence. *European Journal of Behavior Analysis, 10*, 187–202. Retrieved from https://oda.hio.no.

Fields, L., Arntzen, E., & Moksness, M. (2013). Stimulus sorting: A quick and sensitive index of equivalence class formation.

Fields, L., Arntzen, E., Nartey, R. K., & Eilifsen, C. (2012). Effects of a meaningful, a discriminative, and a meaningless stimulus equivalence class formation. *Journal of the Experimental Analysis of Behavior, 97*, 163–181. doi: 10.1901/jeab.2012.97-163

Garner, W. R. (1978). Selective attention to attributes and to stimuli. *Journal of Experimental Psychology, 107*, 287–308. doi: 10.1037//0096-3445.107.3.287

Green, G. (1990). Differences in development of visual and auditory-visual equivalence relations. *American Journal on Mental Retardation*, *95*, 260–270.

Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 229–262). New York, NY: Springer.

Grimm, L. J. (2011). *Emergence of equivalence relations: comparing sorting and match-to-sample procedures.* Nortwestern University.

Holth, P., & Arntzen, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record, 48*, 81–110.

Layng, M. P., & Chase, P. N. (2001). Stimulus-stimulus pairing, matching-to-sample testing, and emergent relations. *The Psychological Record*, *51*, 605–628. Retrieved from http://www.ncbi.nlm.nih.gov/.

Leader, G., & Barnes-Holmes, D. (2001a). Establishing fraction-decimal equivalence using respondent-type training procedure. *The Psychological Record, 51*, 151–165. Retrieved from http://www.ncbi.nlm.nih.gov/.

Leader, G., & Barnes-Holmes, D. (2001b). Matching-to-sample and respondent-type training as methods for producing equivalence relations: Isolating the critical variable. *The Psychological Record, 51*, 429–444. Retrieved from http://www.ncbi.nlm.nih.gov/.

Leader, G., Barnes-Holmes, D., & Smeets, P. M. (2000). Establishing equivalence relations using a respondent-like training procedure III. *The Psychological Record, 50*, 63–78. Retrieved from http://www.ncbi.nlm.nih.gov/.

McIlvane, W. J., & Dube, W. V. (1992). Stimulus control shaping and stimulus control topographies. *The Behavior Analyst, 15*, 89–94. Retrieved from http://www.ncbi.nlm.nih.gov/.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual review of psychology, 32*, 89–115. doi: 10.1146/annurev.ps.32.020181.000513

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology, 115*, 39–57. doi: 10.1037//0096-3445.115.1.39

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology, 13*, 87–108. doi: 10.1037//0278-7393.13.1.87

Pilgrim, C., & Galizio, M. (1996). Stimulus equivalence: A class of correlations or a correlation of classes. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 173–195). Amsterdam: Elsevier.

Rosch, E. (1973). Natural categories. *Cognitive psychology, 4*, 328–350. doi: 10.1016/0010-0285(73)90017-0

Rosch, E. (1999). Principles of categorization. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 189–205). Cambridge, MA: MIT Press.

Rosch, E., & Mervis, C. B. (1975). Families resemblance: Studies in the internal structure of categories. *Cognitive psychology*, *7*, 573–605. doi: 10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Bream, P. (1976). Basic objects in natural categories. *Cognitive psychology, 8*, 382–439. doi: 10.1016/0010-0285(76)90013-X

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

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37*, 5–22. doi: 10.1901/jeab.1982.37-5

Smeets, P. M., Dymond, S., & Barnes-Holmes, D. (2000). Instructions, stimulus equivalence, and stimulus sorting: Effects of sequential testing arrangements and a default option. *The Psychological Record, 50*, 339–354. Retrieved from http://opensiuc.lib.siu.edu/.

Spencer, T. J., & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65*, 643–659. doi: 10.1901/jeab.1996.65-643

Wulfert, E., & Hayes, S. C. (1988). The transfer of conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior, 50*, 125–144. doi: 10.1901/jeab.1988.50-125

Zentall, T. R., Galizio, M., & Critchfield, T. S. (2002). Categorization, concept learning, and behavior analysis: an introduction. *Journal of the Experimental Analysis of Behavior, 78*, 237–248. doi: 10.1901/jeab.2002.78-237

Table 1

Number of Trials and Probability of Programmed Consequences in Training and Testing Blocks

Blocks	Probability of programmed consequences (%)	Minimum trials	Criterion
1. Mixed trials AB, BC, CD, DE	100	60	54/60
2. Mixed trials AB, BC, CD, DE	75	60	54/60
3. Mixed trials AB, BC, CD, DE	25	60	54/60
4. Mixed trials AB, BC, CD, DE	0	60	54/60
Testing			
Test block with *DT trials* (AB, BC, CD, DE),	0	60	54/60
SY trials (BA, CB, DC, ED),	0	60	54/60
TR trials (AC, BD, CE, AD, BE, AE) and	0	90	81/90
EQ trials (CA, DB, EC, DA, EB, EA) randomly intermixed	0	90	81/90

Note. DT = directly trained; SY = symmetry; TR = transitivity; EQ = equivalence.

Table 2

Number of Training Trials and Correct Responses for Each Trial Type in MTS Tests and Sorting Tests

P	Cond	No Trials	DT	SY	TR	EQ	ST
5203	1	360	**60**	**59**	**89**	**89**	**Y**
		480	**57**	**56**	**86**	**86**	**Y**
5207	1	360	**60**	**59**	**89**	**88**	**Y**
		420	**60**	**59**	**85**	**86**	**Y**
5210	1	260	**60**	**60**	**90**	**88**	**Y**
		300	**60**	**60**	**90**	**90**	**Y**
5212	2	480	**59**	**59**	78	80	**Y**
			60	**60**	**89**	**90**	**Y**
5202	2	1080	**58**	**55**	77	74	**Y**
			60	**60**	**90**	**90**	**Y**
5204	4	420	**54**	53	31	32	N
			54	**57**	37	36	**Y**
5211	4	660	**57**	**59**	37	46	N
			56	**55**	37	47	N
5208	4	540	**58**	**57**	55	42	N
			53	**55**	69	58	N
5213	4	1020	**55**	52	27	29	N
			53	52	34	40	N
5214	4	540	**55**	43	31	17	N
			51	47	22	18	N
5201	4	840	50	**55**	30	35	N
			48	50	38	36	N
5206	4	480	41	43	25	30	N
			40	38	27	22	N
5209	4	720	45	37	32	28	N
			22	29	9	12	N
5215	4	1020	30	24	26	24	N
			25	25	35	35	N
5216	4	480	50	28	32	38	N
			45	10	26	33	N
5205	4	960	24	23	27	36	N
			36	23	23	26	N

Note. For each participant the first row show results from Phase 1 and the second row show results from Phase 2. The numbers in bold indicate responding in accordance with experimenter-defined criteria. P = participant; Cond = condition; No Trials = number of training trials; DT = directly trained; SY = symmetry; TR = transitivity; and EQ = equivalence.

Table 3

Pre-class Formation Sorting

P	C-1	C-2	C-3	C-1	C-2	C-3	C-1	C-2	C-3	C-1	C-2	C-3	C-1	C-2	C-3	C-1	C-2	C-3	C-1	C-2	C-3	ECF
5201	**2**	**0**	**1**	**2**	**2**	**1**	**0**	**2**	**1**	**1**	**1**	**2**										**N**
5202	**3**	**1**	**0**	**0**	**1**	**1**	**0**	**2**	**1**	**0**	**1**	**2**	**2**	**0**	**1**							**N**
5203	**0**	**3**	**1**	**3**	**2**	**1**	**2**	**0**	**3**													**Y**
5204	**1**	**2**	**2**	**3**	**1**	**1**	**1**	**2**	**2**													**N**
5205	**0**	**2**	**1**	**0**	**1**	**2**	**3**	**2**	**1**	**2**	**0**	**1**										**N**
5206	**1**	**2**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**2**	**0**	**1**										**N**
5207	**3**	**2**	**1**	**0**	**1**	**3**	**2**	**2**	**1**													**Y**
5208	**2**	**1**	**0**	**1**	**1**	**1**	**0**	**1**	**2**	**0**	**2**	**1**	**2**	**0**	**1**							**N**
5209	**0**	**2**	**1**	**2**	**2**	**1**	**3**	**0**	**2**	**0**	**1**	**1**										**N**
5210	**2**	**1**	**0**	**2**	**0**	**1**	**1**	**1**	**2**	**0**	**2**	**1**	**0**	**1**	**1**							**Y**
5211	**2**	**0**	**1**	**0**	**2**	**1**	**0**	**1**	**2**	**3**	**2**	**1**										**N**
5212	**2**	**2**	**1**	**2**	**2**	**1**	**1**	**1**	**3**													**N**
5213	**4**	**0**	**0**	**0**	**2**	**1**	**0**	**2**	**1**	**1**	**1**	**3**										**N**
5214	**2**	**0**	**1**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**2**	**0**	**1**	**1**	**1**	**1**	**0**	**2**	**1**	**1**	**N**
5215	**4**	**0**	**0**	**0**	**1**	**0**	**1**	**2**	**2**	**1**	**1**	**3**										**N**
5216	**2**	**1**	**1**	**1**	**0**	**1**	**1**	**2**	**0**	**1**	**1**	**1**	**0**	**1**	**2**							**N**

# P	0	0	4	7	4	0	1
# Classes	1	2	3	4	5	6	7

Note. Each column consists of different outcomes dependent on the participants' sorting. Participants sorted correct when he or she had the number 050 meaning zero class-1 members, five class-2 members and zero class-3 members. All together three columns indicates sorting in experimenter-defined classes. The bottom two rows indicate the number of subject which sorted a specific number of classes e.g., four participants sorted the stimuli in three groups. The column to the right shows the outcome on the MTS test for equivalence in Phase 1, and whether it was correct (Y) or incorrect (N). P = participant; c = class; N = no; Y = yes; ECF = equivalence class formation.

Table 4

Post-class Formation Sorting

P	C-1	C-2	C-3	C-1	C-2	C-3	C-1	C-2	C-3	C-1	C-2	C-3	C-1	C-2	C-3	C-1	C-2	C-3
5203	**5**	**0**	**0**	**0**	**5**	**0**	**0**	**0**	**5**									
5207	**5**	**0**	**0**	**0**	**5**	**0**	**0**	**0**	**5**									
5210	**5**	**0**	**0**	**0**	**5**	**0**	**0**	**0**	**5**									
5212	**5**	**0**	**0**	**0**	**5**	**0**	**0**	**0**	**5**									
5202	**5**	**0**	**0**	**0**	**5**	**0**	**0**	**0**	**5**									
5204	**5**	**0**	**0**	**0**	**5**	**0**	**0**	**0**	**5**									
5211	**3**	**0**	**0**	**2**	**3**	**0**	**0**	**2**	**5**									
5208	**5**	**0**	**0**	**0**	**5**	**0**	**0**	**0**	**5**									
5213	**2**	**0**	**1**	**2**	**1**	**0**	**1**	**2**	**1**	**0**	**1**	**2**	**0**	**1**	**1**			
5214	**3**	**2**	**3**	**2**	**3**	**2**												
5201	**2**	**2**	**0**	**0**	**1**	**5**	**3**	**2**	**0**									
5206	**3**	**0**	**0**	**2**	**2**	**1**	**0**	**1**	**2**	**0**	**2**	**2**						
5209	**3**	**1**	**3**	**0**	**2**	**1**	**2**	**2**	**1**									
5215	**4**	**1**	**0**	**0**	**2**	**3**	**1**	**2**	**2**									
5216	**1**	**1**	**1**	**2**	**2**	**1**	**2**	**2**	**3**									
5205	**3**	**1**	**1**	**0**	**1**	**2**	**0**	**2**	**0**	**2**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**2**

# P	0	0	8	4	1	2
# Classes	1	2	3	4	5	6

Note. Each column consists of different outcomes dependent on the participants' sorting. Participants sorted correct when he or she had the number 050 meaning zero class-1 members, five class-2 members and zero class-3 members. All together three columns indicates sorting in experimenter-defined classes. The bottom two rows indicate the number of subject which sorted a specific number of classes e.g., eight participants sorted the stimuli in three groups. P = participant; c = class.

Table 5

Postpost-class Formation Sorting

P	C-1	C-2	C-3	C-1	C-2	C-3	C-1	C-2	C-3	C-1	C-2	C-3	C-1	C-2	C-3	C-1	C-2	C-3
5203	5	0	0	0	5	0	0	0	5									
5207	5	0	0	0	5	0	0	0	5									
5210	5	0	0	0	5	0	0	0	5									
5212	5	0	0	0	5	0	0	0	5									
5202	5	0	0	0	5	0	0	0	5									
5204	5	0	0	0	5	0	0	0	5									
5211	3	0	0	2	3	0	0	2	5									
5208	5	0	0	0	5	0	0	0	5									
5213	2	0	1	2	1	0	1	2	1	0	1	2	0	1	1			
5214	3	2	3	2	3	2												
5201	2	2	0	0	1	5	3	2	0									
5206	3	0	0	2	2	1	0	1	2	0	2	2						
5209	3	1	3	0	2	1	2	2	1									
5215	4	1	0	0	2	3	1	2	2									
5216	1	1	1	2	2	1	2	2	3									
5205	3	1	1	0	1	2	0	2	0	2	0	0	0	1	0	0	0	2

# P	0	1	12	1	1	1
# Classes	1	2	3	4	5	6

Note. Each column consists of different outcomes dependent on the participants' sorting. Participants sorted correct when he or she had the number 050 meaning zero class-1 members, five class-2 members and zero class-3 members. All together three columns indicates sorting in experimenter-defined classes. The bottom two rows indicate the number of subject which sorted a specific number of classes e.g., 12 participants sorted the stimuli in three groups. P = participant; c = class.

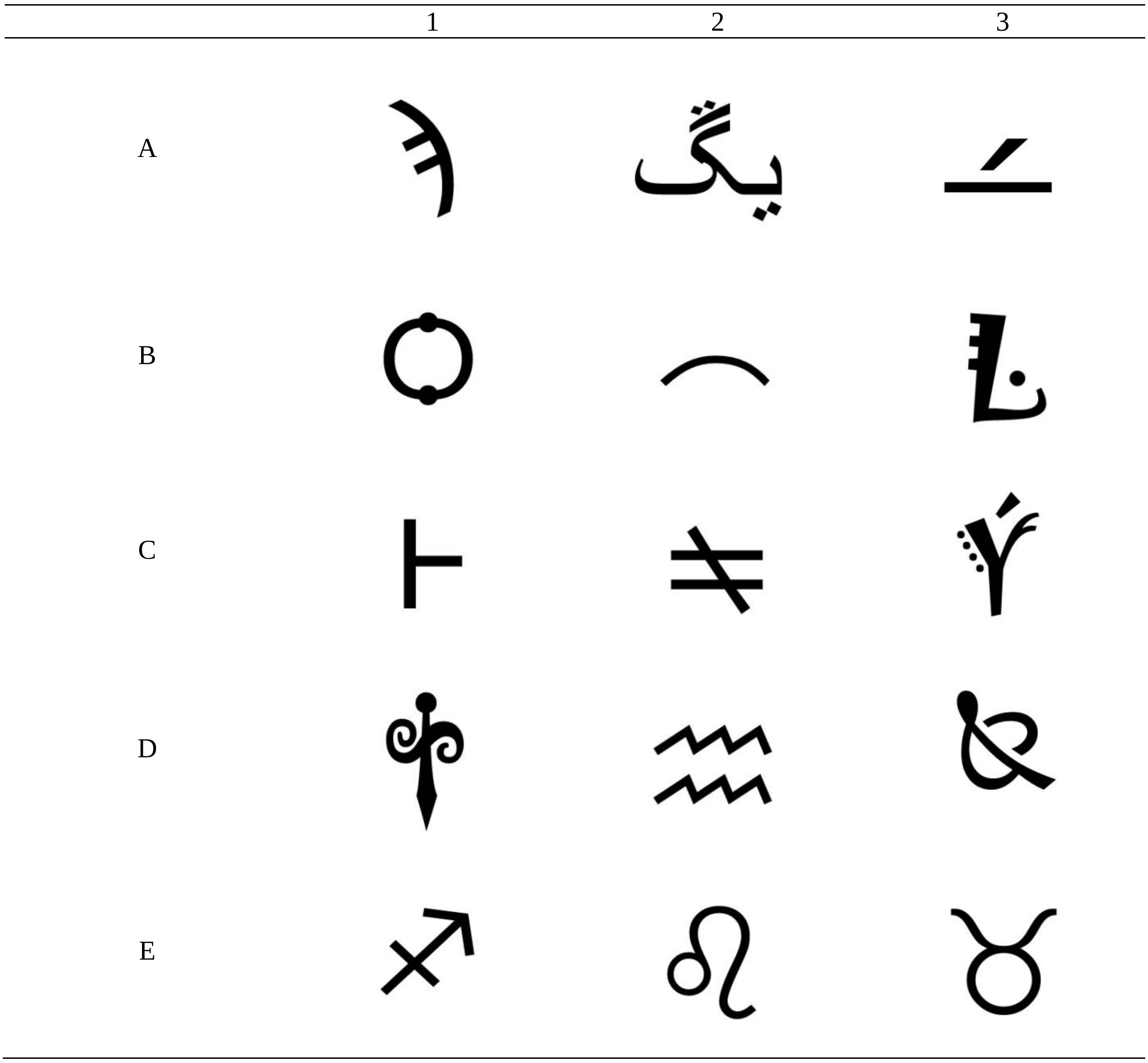

Figure 1. Stimulus set 1 used in Phase 1 in which every participant conducted. The columns (1-3) represent the classes and the rows (A-E) represent the members of the class e.g. A1/B1/C1/D1/E1 is in the same experimenter-defined class.

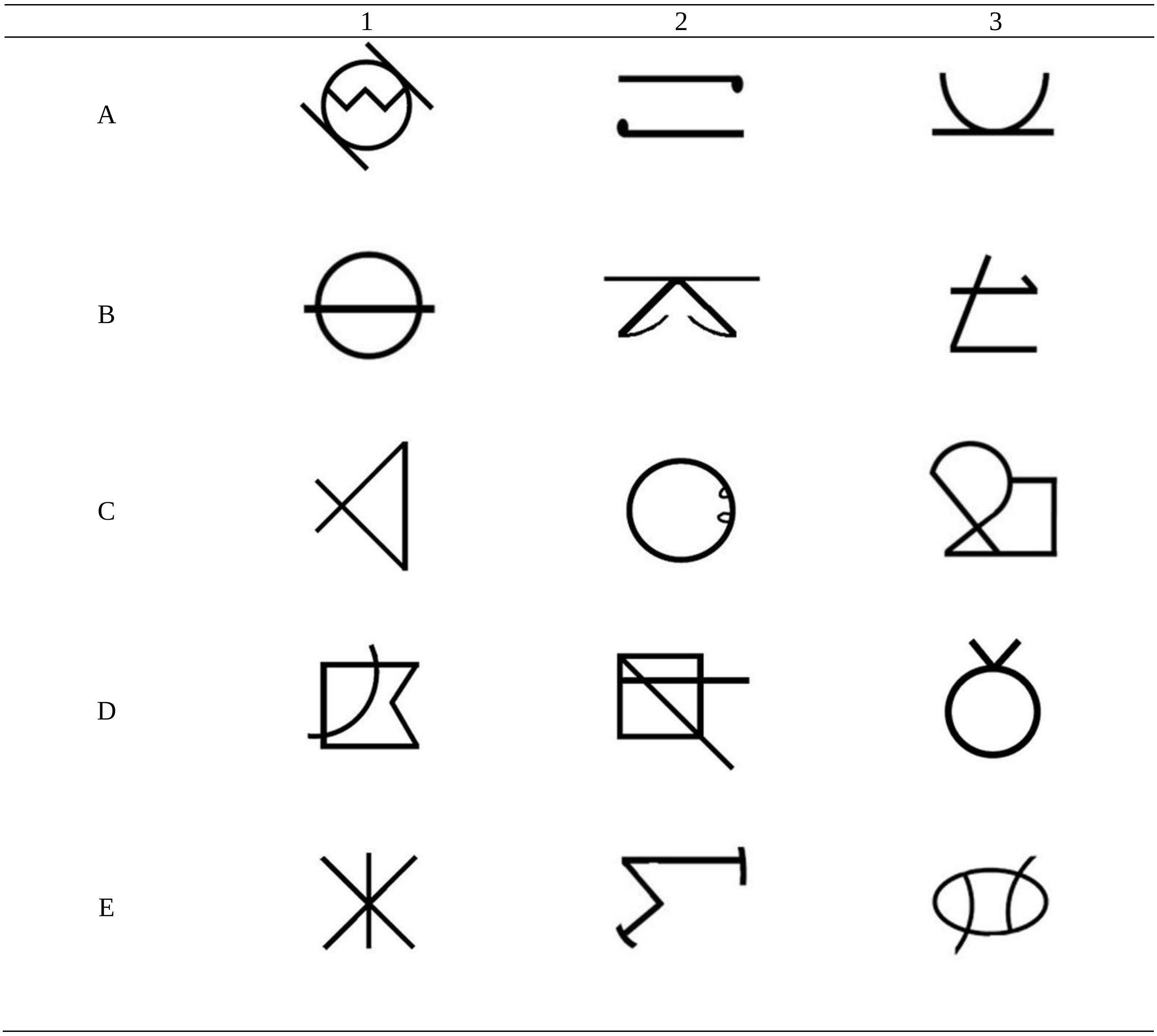

Figure 2. Stimulus set 2 used in Phase 2, Condition 1 in which three participants conducted. The columns (1-3) represent the classes and the rows (A-E) represent the members of the class e.g. A1/B1/C1/D1/E1 is in the same experimenter-defined class.

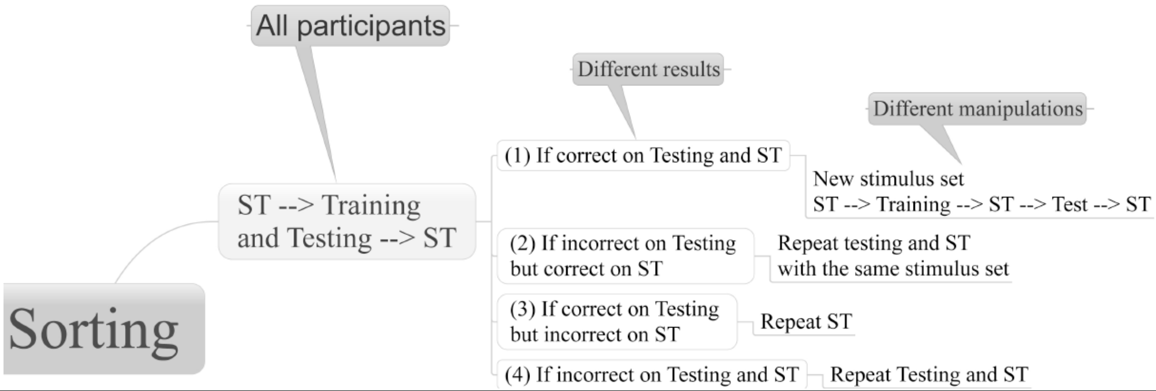

Figure 3. All participants went through Phase 1 (first column) containing a sorting test, matching-to-sample training and testing, and a new sorting test. Dependent on the results from Phase 1, participants were allocated to one of four different conditions (second and third column). ST = sorting test.

Symmetry	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3
B1	**5**											
B2		**5**										
B3		2	3									
C1				2	3							
C2					**5**							
C3						**5**						
D1							**5**					
D2								**5**				
D3									**5**			
E1										**5**		
E2											**5**	
E3												**5**

Equivalence	A1	A2	A3	B1	B2	B3	C1	C2	C3
C1	**5**								
C2	4		1						
C3		1	4						
D1		5		4	1				
D2	2		3		**5**				
D3		1	4	1		4			
E1	3	1	1	3	2		**5**		
E2	3	2		2	3			**5**	
E3	5			3	2		5		

Figure 4. Matrix for P5211 displaying results from MTS- test in Phase 2.

Symmetry	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3
B1			5									
B2	5											
B3		5										
C1					4	1						
C2				5								
C3						5						
D1							5					
D2								5				
D3									5			
E1										5		
E2												5
E3											5	

Equivalence	A1	A2	A3	B1	B2	B3	C1	C2	C3
C1	5								
C2			5						
C3		5							
D1	5				5				
D2			5	5					
D3		5				5			
E1	5				5		5		
E2		5				5			5
E3			5	5				5	

Figure 5. Matrix for P5215 displaying results from MTS-test in Phase 2.

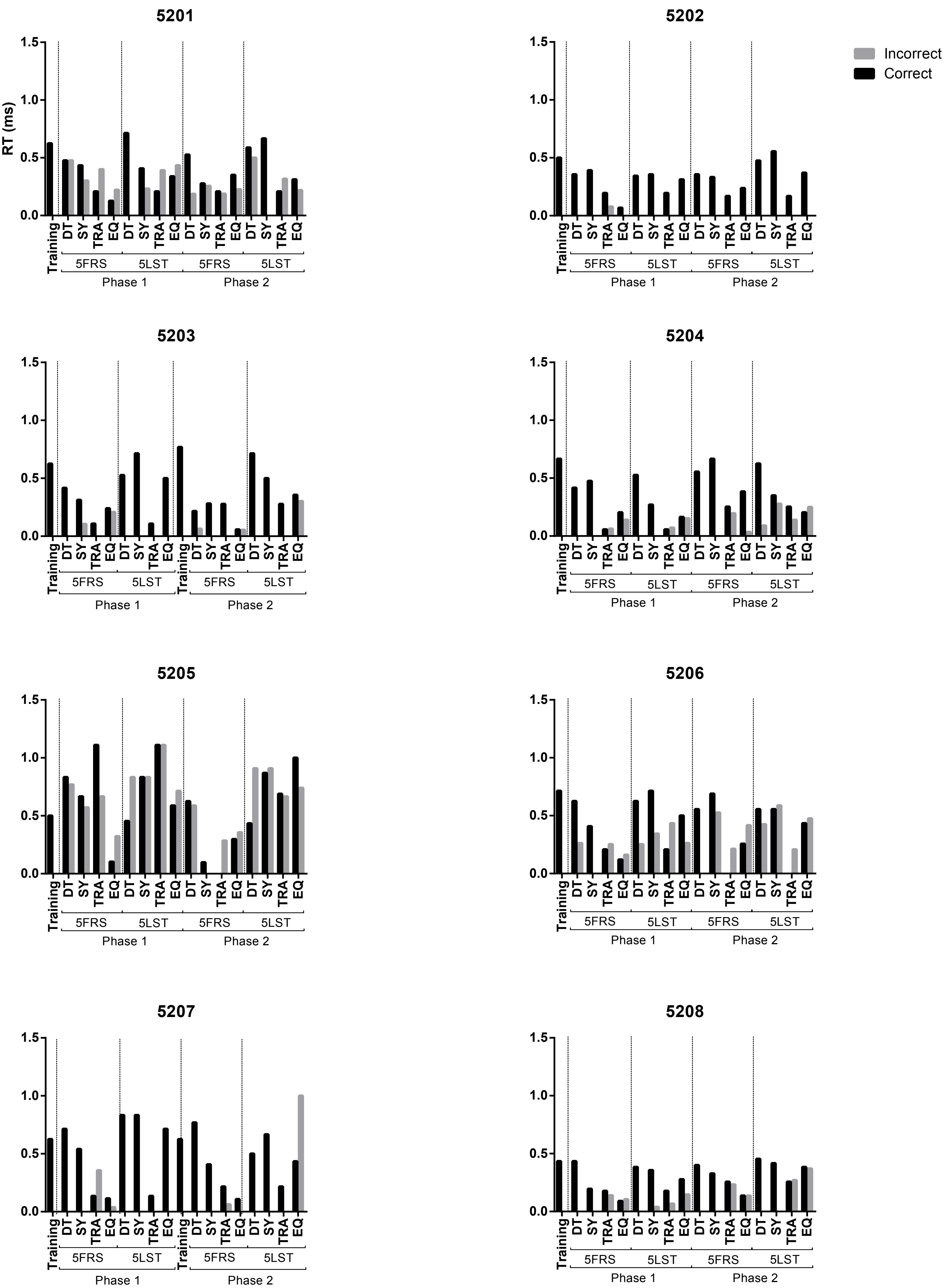

Figure 7. Showing inversed reaction time for P5201-P5208. Grey bars indicating incorrect responses and black bars indicating correct responses. Data taken from the five first and five last trials in the MTS-test from each phase. DT = directly trained, SY = symmetry, TR = transitivity, EQ = equivalence; FRS = first; LST; last.

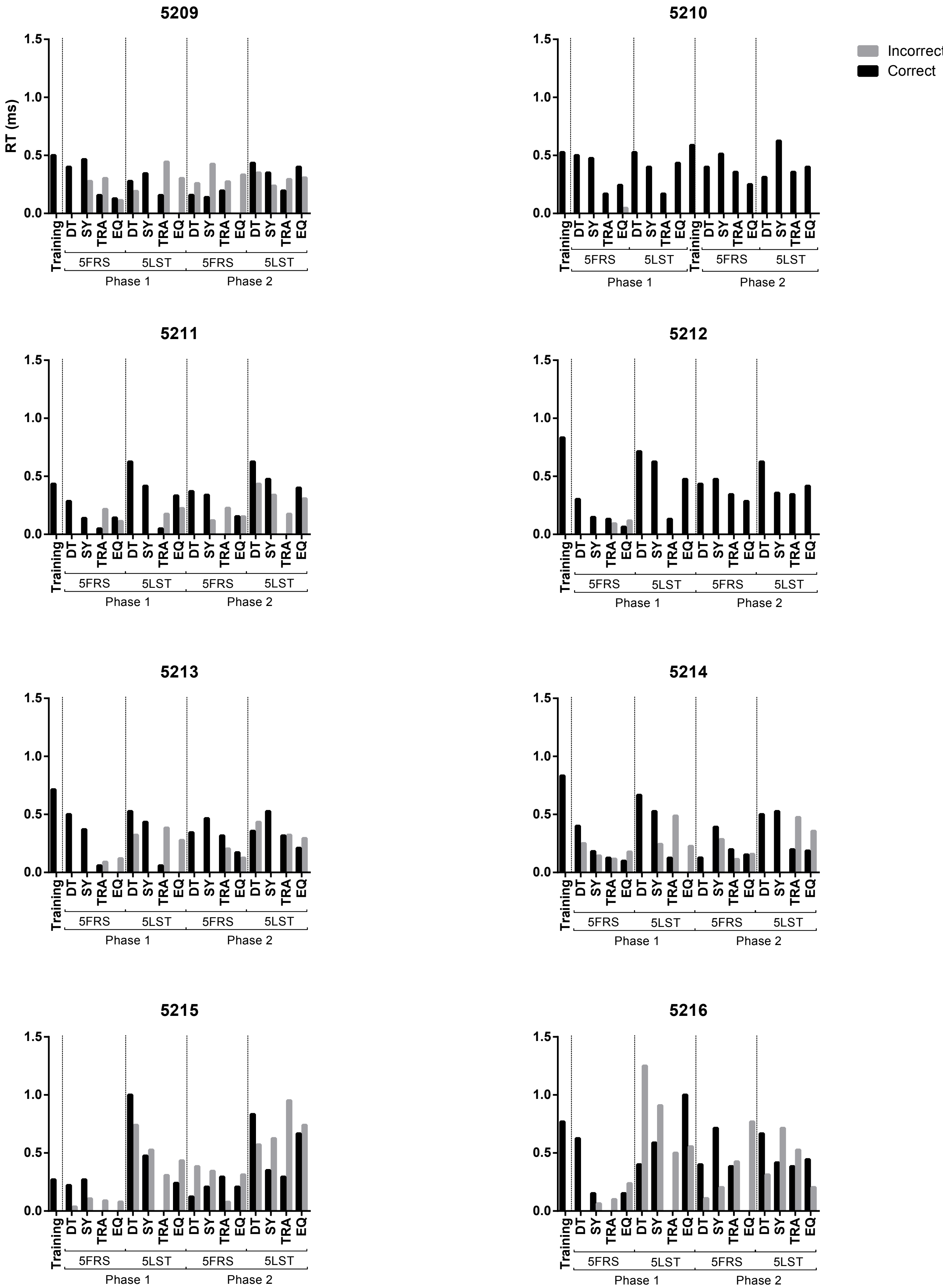

Figure 7. Showing inversed reaction time for P5209-P5216. Grey bars indicating incorrect responses and black bars indicating correct responses. Data taken from the five first and five last trials in the MTS-test from each phase. DT = directly trained, SY = symmetry, TR = transitivity, EQ = equivalence; FRS = first; LST; last.