MASTER THESIS
Learning in complex systems
May 2017

Stimulus Equivalence Class Formation and Stimulus Sorting

Eva Lyholm Limi

Faculty of Health Sciences
Department of Behavioral Science

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Contact:

Eva Lyholm Limi

Luksefjellvegen 276

3721 Skien

Thanks to Erik Arntzen for supervision, Aleksander Vie for help with data programs, Richard K. Nartey for help with statistic and to my husband and family for their patience with me during my work on this paper.

Table of Contents

Theoretical Part

Class Formation Measured by Sorting: A Review

Table of contents

Experimental Part

Attending Behavior in the Presence of Compound Stimuli

Class Formation Measured by Sorting: A Review

Abstract

Historically, sorting tests have been used as a secondary or an additional measure for stimulus class formation. In these early studies, sorting tests were administered subsequently to matching-to-sample (MTS) tests for stimulus equivalence. Recently, the immediate emergence of equivalence classes has been documented in sorting tests in two experiments. Discordant findings — showing that classes documented by sorting tests immediately after the training of baseline relations cannot always be defined as equivalence classes — have also been identified. To investigate these types of discordant findings further, we reviewed 25 articles. The reviewed articles were selected from searches in PsycINFO and by examining the reference lists in the already identified articles. The inclusion criterion was whether sorting tests were used in relation to MTS-based training or a similar procedure to establish conditional discrimination. Studies that met the criterion were quantified, evaluated and discussed along a number of dimensions related to parameters and variables in sorting procedures and in MTS training procedures.

Keywords: sorting, stimulus equivalence, class formation, immediate emergence

Through Sidman's analysis of stimulus equivalence (Sidman, 1971, 1994; Sidman & Tailby, 1982), it has been possible to investigate complex human behavior, which is colloquially called language, symbol use, memory and problem solving. These investigations are done by studying the variables that affect how categorizing emerges and those influencing the formation of stimulus classes. Sidman and Tailby (1982) defined the conditional relations between the stimuli as *stimulus equivalence* if the participants respond with interchangeability between the stimuli with respect to intact reflexivity, symmetry and transitivity relations. Sidman (1994) explained that reflexivity relations can be inferred when all the involved stimuli are demonstrated to be related to themselves. The symmetry relations are shown when sample-comparison interchangeability is demonstrated and transitivity relations are shown by demonstrating conditional relations between stimuli that have not been directly related to each other in training but are related to one another by the conditional relation to another stimulus. Stimulus equivalence can be illustrated by the following example for a person who has no knowledge of dogs: three arbitrary stimuli could be the sound of the word *dog* (A); a picture of a dog (B); the printed word *DOG* (C) and if the sound of the word *dog* is trained to the picture of a dog (A –> B) and the picture of a dog is trained to the printed word *DOG* (B –> C). We would then test whether the untrained conditional relations emerge, being the relations of symmetry (B–>A) and (C–>B), transitivity (A–>C) and global equivalence (C –>A), and if the participant also matches A to A, B to B and C to C. We would say that an equivalence class with three members has emerged.

Sidman and Tailby (1982) described the MTS procedure used for training and testing conditional relations and stimulus equivalence with non-identical, arbitrary or symbolic stimuli. To describe the MTS procedures, they identified the stimuli with an alphanumeric code, naming the classes using a number — Class 1, Class 2 and so forth — and naming its members using capital letters. Class 1 with three members then consists of the members A1, B1 and C1; Class 2 consists of A2, B2 and C2 and so forth. When forming three 3-member classes, the stimulus set of members used in training consists of all the As (i.e., A1, A2, and A3), all the Bs (i.e., B1, B2, and B3) and all the Cs (i.e., C1, C2 and C3). An MTS trial often involves the following: an observing response to the sample stimulus (i.e., the conditional sample stimulus) is given by the participant and is then followed by the presentation of comparison stimuli (i.e., the discriminative events). One of the comparison stimuli (i.e., the experimenter-defined stimuli) matches the sample stimuli, and the selection of this will produce reinforcement. Choosing a comparison that does not match the sample will not be reinforced. After training the conditional discrimination (i.e., the baseline relations), a test would normally be included testing for the emergence of new relations and whether the baseline relations are maintained when they are presented together with the new trials (Sidman, 1994).

There are three training structures in which MTS-based training can be conducted and the emergent relations tested afterwards will vary based on the test used. One is the *linear series* (LS) training structure in which first all AB and then all BC relations would be trained for three 3-member classes. Subsequent testing determines whether the trained relations are maintained as well as whether the BA, CB, AC and CA relations have emerged. Another is the *many-to-one* (MTO) training structure; in this the AC and BC relations are trained. In the test afterwards is tested for the maintenance of baseline relations and the emergent relations (i.e., CA, CB, AB and BA). Last is the *one-to-many* (OTM) training structure in which the AB and AC relations are trained and subsequent baseline relations and the emergent relations (i.e., BA, CA, BC and CB) are tested.

Another way in which the training and testing structure can vary is in three different protocols. In the simple-to-complex protocol, all baseline relations are trained, and symmetry, transitivity and equivalence tests are interspersed incrementally before a mixed test that includes all relations is performed. In the complex-to-simple protocol, all baseline relations are trained, and the equivalence relations are tested before a mixed test that includes all relations is conducted. In the simultaneous protocol (in which most of the equivalence research is done (Arntzen, 2012a)), training of all baseline relations is done before testing for any of the emergent relations (Imam, 2006).

Other variables and parameters in MTS-based research procedures can vary — for example, the number of classes and members, the arrangement of training trials (e.g., gradual introduction of training trials or not, serialized or concurrent presentations of trials) and whether the procedure involves a simultaneous matching to sample or delayed matching to sample. Additionally, the stimuli used can vary, for example, between abstract, non-figurative, nonsense syllables and pictures (Arntzen, 2012b).

The MTS-based test has been commonly used in the field of stimulus equivalence, whereas sorting tasks have been applied in many areas of psychology over a long period of time — for example, when documenting categorization and concept formation (Ludvigson & Caul, 1964; Rosch & Mervis, 1977), early investigations in the field of behavioral phenomena often had their conclusions based upon the means of the measured behavior throughout the experiment (Dymond & Rehfeldt, 2001). Dymond and Rehfeldt (2001) believed that it would be an advantage to expand the methods used for measuring the emergence of untrained conditional discriminations. They argued that additional measures such as reaction time, verbal reports, stimulus recall, and stimulus sorting could allow for convergent validity for the means of measured behavior and could provide more knowledge to the variables responsible for derived responding. Sorting has in some studies, been used to track the formation and maintenance of equivalence classes, and sorting has been studied to ascertain whether the results of the sorting tests are reliable with respect to the emergence of equivalence classes. The main purposes of this article are to reveal and to discuss research in which sorting is used as a measure for emergence or maintenance of equivalence classes and to reveal and discuss the research on sorting tests. In both cases, we will look into different parameters used in the sorting tests and in the related conditional discrimination training procedures, which are mostly MTS-based training and testing procedures. The yields of the sorting tests and outcome of MTS tests are discussed with respect to whether the difference in parameters used in training and testing can influence the function of the classes documented by sorting tests.

Method

Studies in stimulus equivalence using sorting as an additional measure for class formation and studies in which sorting as an alternative measure for class formation were investigated were identified through advanced searches in PsycINFO, which were conveyed through the Ovid® technologies databases, which are part of the Wolters Kluwer group. The most recent search was done on the 29th of April 2017 using the keyword *sorting tests.* The keywords were combined with "and," and the search was limited to *humans.* The identified articles were reviewed to determine inclusion in the present review. The reference lists of the identified articles were examined to find additional articles. Only articles found using this method were included in the review. We utilized PsycINFO for searching because it is a highly recognized database within psychology and the behavioral and social sciences.

Inclusion Criteria

The articles included were those in which either a sorting post-class formation test or a sorting pre-class formation test and a sorting post-class formation test together were used in relation to an MTS-based or a similar procedure to train and test the formation of the equivalence classes. Articles studying sorting tests as an alternative measure of class formation were also included.

Data Display and Analysis

Studies that were included in the present review varied along several dimensions related to the participants and setting characteristics, parametric characteristics of the sorting test and MTS training and testing and varied in the distribution of classes and how stimuli were arranged. To be able to quantify and evaluate along the variation of dimensions, a matrix (Table 1) with parameters and findings was made. In Table 1, the articles were listed vertically in alphabetical order. The characteristics, parameters, and the findings on sorting are presented in the rows for each study under the following headlines: *Authors* (under which the authors of the article are listed), *Year* (the year the article is published), *Journal* (the journal in which the article is published), *Number of participants/ Population details /Age in years* (the details about the population used), *Numbers of classes and members and stimuli used* (the type of stimuli used and how the setups of stimuli classes were organized), *Some parameters from the procedures of sorting* (if available, the instruction used in the experiment before the sorting test, and how the sorting procedure was conducted), *Some parameters from the procedures of training of baseline relations* (the parameters from the MTS training that were related to the sorting test), *Some parameters from the procedures of testing equivalence* (the parameters from the MTS test that were related to the sorting test), and *Quotes and comments about the findings and conclusions on sorting* (the findings on sorting were presented by quotes from the articles; when no single quote was able to summarize the findings, the findings were presented as a comment).

The selected articles were reviewed and quantified or summarized under the following headings: *Procedural Variables for Both MTS and Sorting* (i.e., the variables being the same for both type of tests are reviewed), *Differences in the MTS Training* (i.e., only different training structures are included in this review), *Differences in the Sorting Procedures, Sorting as Post-Class Formation Tests, Sorting as Pre- and Post-Class Formation Tests, Concordant and Discordant Findings, Maintenance of Stimulus Classes, Delayed Emergence, Research* *on Sorting,* and *Economy of Sorting Tests.* The selections of the articles summarized under a certain headline were made because these particular articles were interesting examples that illustrate the theme.

Results

Using the search strategy described above, the search yielded a total of 56 published works. Of these, 50 articles were excluded from the analysis because sorting tests were not used in relation to a matching-to-sample procedure. The articles found and included were marked with an asterisk in the *Authors* column in the matrix (see Table 1). By examining the reference lists of the six articles, a total of 24 articles and one chapter of a book was included in the analysis. As seen in Table 2, we found that *The Psychological Record* published the majority (44%) of the reviewed studies; the *European Journal of Behavior Analysis* published 20%; and the *Journal of the Experimental Analysis of Behavior* published 20%. The remainder of the included studies were either published by the *Journal of Applied Behavior Analysis, American Journal on Mental Retardation, Learning Behavior* or as a chapter in a book.

Reliability

The exact same procedure to search for articles in PsycINFO was used by a second reader, who found the exact same six articles, resulting in an inter observer agreement (IOA) of 100%.

Procedural Variables for Both MTS and Sorting

Population characteristics. Most studies (92%) included adults, and only 2 studies (8%) included children (see Table 3). At the same time, the majority of the studies were conducted with students at different levels of education (60%). Only 8% of the studies included participants with mental disabilities. In 32% of the studies, the details of the participants were not specified beyond stating whether the participants were children or adults.

Stimuli. When we reviewed the articles focusing on the stimuli used, we found several differences in the types of stimuli used between the studies. In Table 4, the types of stimuli used in the reviewed articles are summarized. We see that 28% of the studies were conducted with abstract stimuli printed in black on a white background (Arntzen, Braaten, Lian, & Eilifsen, 2011; Arntzen, Granmo, & Fields, 2017; Arntzen, Norbom, & Fields, 2015; Eilifsen & Arntzen, 2009, 2011; Fields, Arntzen, & Moksness, 2014; Mackay, Wilkinson, Farrell, & Serna, 2011). Additionally, 16% of the studies used abstract stimuli with meaningful stimuli (Arntzen, Nartey, & Fields, 2014; Fields, Arntzen, Nartey, & Eilifsen, 2012; Nartey, Arntzen, & Fields, 2014, 2015). The remainder of the 14 studies used stimuli that differed in several dimensions — for example, abstract objects (Pilgrim & Galizio, 1996), Icelandic spoken nouns (Sigurdarddottir, Mackay, & Green, 2012) and pictures of students (Dickins, 2011).

Number of stimulus classes and members. In Table 5, we see the number of classes and numbers of members within the classes represented in the reviewed articles. In 44% of the articles, three 5-member classes were used. In 16% of the studies, three 3-member classes were used, and in 12% of the studies, two 3-member classes were used. The remainder of the studies used two 4-member classes, two 5-member classes, two 7-member classes, three 4-member classes, four 3-member classes or six 3-member classes.

Procedural Differences in the MTS Training

Training structure. Different procedures in the MTS training procedures were found. For example, an LS training structure in several studies (Arntzen et al., 2011; Arntzen et al., 2017; Arntzen et al., 2014; Arntzen et al., 2015; Eilifsen & Arntzen, 2009, 2011; Fields et al., 2014; Fields et al., 2012; Mackay et al., 2011; Nartey et al., 2014, 2015; Nedelcu, Fields, & Arntzen, 2015; Pilgrim & Galizio, 1996; Smeets, Dymond, & Barnes-Holmes, 2000) was used, whereas an MTO training structure was used in three studies (Arntzen, 2004; Hove, 2003; Varelas & Fields, 2017) and an OTM training structure was used in several studies (Fienup & Dixon, 2006; Green, 1990; Hove, 2003; Sigurdarddottir et al., 2012; Smeets & Barnes-Holmes, 2005). For other studies, training structure was not relevant or not specified (Cowley, Green, & Braunling-McMorrow, 1992; Dickins, 2011, 2015; Lowe, Horne, Harris, & Randle, 2002).

Procedural Differences in Sorting Tests

Conducting the sorting test. Most studies (68%) used a tabletop procedure with cards to conduct the sorting test (see Table 6). Other studies (20%) used a computer to conduct the sorting test. Three studies had different procedures — for example, asking the child to give Teddy the other objects (Lowe et al., 2002), using a pencil and paper procedure (Smeets et al., 2000) or grouping objects on a table (Pilgrim & Galizio, 1996).

Instructions. When we examined the studies in this review (see Table 7), we found several different instructions that were used before the sorting tasks. An instruction that included the words "put them into groups" was the most common instruction used by seven of the 25 reviewed articles (Arntzen et al., 2014; Dickins, 2011; Fields et al., 2014; Fields et al., 2012; Nartey et al., 2014, 2015; Pilgrim & Galizio, 1996). In four studies, the participants were told to categorize the stimuli (Arntzen, 2004; Arntzen et al., 2011; Eilifsen & Arntzen, 2009, 2011), and two studies used instructions to sort the cards into stacks (Green, 1990; Nedelcu et al., 2015). In other studies, a combination of instructions was used, for example, using both the words "categorize" and "into groups" (Smeets et al., 2000); "that go together" and "in stacks" or "piles" were also used in combination (Arntzen, 2004; Cowley et al., 1992). The remainder of the studies used different instructions, or what was said to the participants prior to the sorting test was not specified in the text.

Organization of stimuli in sorting tests. The procedure of the sorting test was conducted in different ways in relation to how the participants were supposed to organize the stimuli. In 12 studies, the participants made clusters or groups or arranged the cards, and the stimuli were removed from a stack/deck of cards and placed on the table or on the computer screen in a way that made all the stimuli visible at the same time (Arntzen et al., 2017; Arntzen et al., 2014; Arntzen et al., 2015; Dickins, 2011, 2015; Fields et al., 2014; Fields et al., 2012; Nartey et al., 2014, 2015; Pilgrim & Galizio, 1996; Sigurdarddottir et al., 2012; Smeets & Barnes-Holmes, 2005; Varelas & Fields, 2017). In five studies, the participants placed the cards or objects into stacks or piles (Cowley et al., 1992; Fienup & Dixon, 2006; Green, 1990; Nedelcu et al., 2015; Smeets et al., 2000). When sorting the cards and placing them into stacks, that would make only the upper card in the stacks visible, but the studies did not clarify whether the stacks were locked or whether the participant could scroll through the stacks. In the following eight studies, it was not specified how the sorting was conducted in the sense of clusters vs. stacks (Arntzen, 2004; Arntzen et al., 2011; Eilifsen & Arntzen, 2009, 2011; Hove, 2003; Lowe et al., 2002; Mackay et al., 2011; Sigurdarddottir et al., 2012).

Examples of early investigation with sorting tests. Green (1990), for example, investigated the differences in developing equivalence in visual and auditory-visual relations. In one condition, Green used abstract symbols. In another condition, Green used a mix of abstract symbols with nonsense syllables. The participants were five young adult women with mild retardation. The training was conducted on paper. The sorting test was used to provide more information about the classes of stimuli that had emerged through the MTS training and testing under the two conditions. The test was conducted with cards. Green (1990) concluded that the results of the sorting tests were consistent with the other findings; the auditory-visual classes emerged more quickly and were also sorted immediately by all participants, whereas the visual classes did not emerge as quickly, and only two participants managed to sort the stimuli correctly into the experimenter-defined classes.

In a replication done by Smeets et al. (2000), with six symbols as experimental stimuli, it was found that the performance on sorting tests was related to the number of derived stimulus relations. Smeets et al. (2000) found that 61% of the participants who did not meet the criteria for stimulus equivalence nevertheless sorted the stimuli into experimenter-defined classes. Therefore, they concluded that the sorting test used in the experiment did not provide convergent validity for stimulus equivalence.

Different procedures for sorting were found. For example, in the study on naming and categorization in children in Lowe et al. (2002), they discuss the use of a sorting test that they referred to as *category match to sample.* In Lowe et al., children were first taught to tact two arbitrary and abstractly shaped sets of wooden stimuli, and in a sorting test afterwards, they were tested in terms of categorizing the arbitrary stimuli. The children were asked to look at the sample and find the others of same type, in the following way: “Look at this. Can you give Teddy the others?” If the child selected all the comparison stimuli, they would get the following instruction: “Teddy doesn´t want all of them, only some,” and the trial would be repeated and the child asked to tact the stimulus prior to selecting the other stimuli.

In the Mackay et al. (2011) study on merging and the intersection of equivalence classes, we found slightly different procedures for sorting. Mackay et al. (2011) used a sorting/matching method they called *multi-selection matching-to-sample* that was arranged on the computer in the following way: a sample stimulus was shown on the screen. When the participant had looked at and clicked on the stimulus, three other stimuli appeared on the screen. The participants were instructed to select *all* the stimuli that were equivalent to the sample stimulus and to pick as many pictures as they thought would fit.

Sorting as Post-Class Formation Tests

In the early studies using sorting in combination with the MTS-based test for stimulus equivalence, the sorting tests are primarily used as an additional post test for class formation. (e.g., Arntzen, 2004; Dickins, 2011; Fienup & Dixon, 2006; Green, 1990; Hove, 2003; Lowe et al., 2002; Mackay et al., 2011; Pilgrim & Galizio, 1996; Sigurdarddottir et al., 2012; Smeets et al., 2000).

The study of Hove (2003) is an example of a study using sorting as a post-class formation test. Hove investigated the probability of obtaining an equivalence class formation after a one-to-many (OTM) training structure and after a many-to-one (MTO) training structure. He conducted the experiment with 20 college students. One group underwent the training and testing in a MTS procedure with an MTO structure, and the other group underwent the training and testing in an OTM structure. Greek letters were used as stimuli, and the experiment was conducted on a personal computer. The sorting tests were used as post-experimental sorting tasks with cards. The participants "were asked to sort the nine cards"(Hove, 2003, p. 621). Hove found that six out of the ten participants in the OTM group sorted the cards according to the experimenter-defined classes. Of these six participants, four did not respond according to equivalence on the prior MTS-based test for emergent relations. Ten out of ten participants sorted the cards according to experimenter-defined classes in the MTO group. Of these ten participants, two did not respond according to equivalence on the prior MTS-based test.

Arntzen (2004) used sorting as a post-class formation test. The probability of stimulus equivalence formation was investigated when familiar pictures were used with abstract stimuli in an MTS procedure and in different positions in the MTO-training structure. At the end of the study, the participants were given printouts of the stimuli and were asked "to categorize the stimuli." The experiment showed that participants who responded according to stimulus equivalence also categorized the printout stimuli in the three experimenter-defined groups, while the participants who did not respond according to stimulus equivalence did not sort the stimuli according to the experimenter-defined groups.

A sorting post-class formation test is also used in Fienup and Dixon (2006). In this study, they investigated a cross-modal equivalence formation between three visual-visual classes with three members and three visual-olfactory classes with three members. The visual stimuli used were pictures of patterns of, for example, rugs and sweaters that were printed on cards and laminated; the olfactory stimuli were kept in plastic squeeze bottles of equal size. The sorting test was performed last in the experiment after training and testing to merge the classes. The participants were given all the stimuli used in the experiment and were asked "to place the objects in three piles" (Fienup & Dixon, 2006, p. 92). It was found that all participants met the criteria for demonstrating equivalence in the sorting test, whereas only one participant before this test had demonstrated equivalence in the merge test. The experimenters explain the discrepancy as being due to delayed emergence of derived relations.

Sorting as Pre- and Post-Class Formation Tests

Other studies involved sorting tests as both a pre-class formation test to test for entry-level skills and as an additional post-class formation test for class formation to use the changes from the pre-sorting to the post-sorting procedure to assess class formation (e.g., Arntzen et al., 2011; Arntzen et al., 2017; Arntzen et al., 2015; Cowley et al., 1992; Eilifsen & Arntzen, 2009, 2011; Fields et al., 2014; Fields et al., 2012).

One example of an early study using pre-class formation sorting tests is Cowley et al. (1992). A pre-class formation sorting test together with naming tests to reveal the participants' entry-level skills was used. Matching-to-sample procedures were used to teach the three adults with brain injuries name and face matching. All three participants demonstrated the formation of three equivalence classes, each including a therapist's dictated name, photo and written name in a matching-to-sample procedure. In the post-class formation sorting test, the participants were instructed to “put all the cards that go together into stacks” (p. 466). One participant was discharged from the institution before completing the experiment, and two participant grouped all stimuli correctly.

Fields et al. (2012) used sorting as pre-class and post-class formation tests. The study was conducted to investigate the effects on equivalence class formation related to three different conditions with ten participants in each. The training was conducted on a computer in the MTS format and the participants were trained to form three 5-member classes with concurrent training and in a linear series training structure. It was found that when all stimuli were abstract, none of the participants formed classes. When the C stimuli were a meaningful stimuli and the remainder of stimuli were abstract, eight of the ten participants formed classes in the emergent relations test blocks. It was also found that when the abstract C stimuli became an S^D before class formation, in a mix of simultaneous and successive discrimination training, five out of ten participants formed classes. Sorting tests were used as a pre-class formation test prior to the training of baseline relations and showed that no participants spontaneously sorted the stimuli into the experimenter-defined classes. After the emergent relations MTS test blocks, a new sorting test was conducted as a post-class formation test. The sorting tests were conducted with plastic-laminated cards and included the following instruction: “Please put the cards into groups and call me when you have completed the task” (p.168). Thus far, the sorting procedures have been similar to other procedures we have mentioned, but Fields et al. (2012) took the analysis of the sorting results a step further by analyzing every single class formation cluster the participants sorted to see whether it was possible to identify any consistent participant-defined classes between the emergent test performances and the sorting tests. When Fields et al. (2012) analyzed the emergent relations MTS test performances trial-by-trial. It was found that “the effects of the C stimulus functions on class formation were the same when measured by sorting and the derived relations test" (p. 171). For four participants, dissociation was seen as they documented Class 1 by the card sorting tests but did not with the emergent relations tests. Fields et al. (2012) inferred from these results that "the card-sorting test may be a more sensitive measure of class formation than the emergent relations test. The card sort was a second measure of class formation and may have tracked the first stages of the delayed emergence of all three classes" (p. 173).

Concordant and Discordant Findings

In many experiments in recent years, the findings on the post-class formation sorting test are clearly concordant with the findings on the MTS-based emergence of equivalence classes test (e.g., Arntzen, 2004; Arntzen et al., 2014; Fields et al., 2014; Nartey et al., 2014, 2015; Nedelcu et al., 2015; Travis, Fields, & Arntzen, 2014).

Nartey et al. (2015), for example, found concordance between the MTS-based equivalence tests and the post-class formation sorting test when they replicated the study of Fields et al. (2012). For the results on sorting, this study found the same results as in Fields et al. (2012). None of the participants sorted the stimuli according to the experimenter-defined classes in the pre-sorting test; however, after training and testing, all participants who formed the experimenter-defined classes in the test for emergence also sorted the stimuli into the classes according to equivalence. Additionally, the participants who did not form the experimenter-defined classes in the test for emergence did not sort the stimuli into the classes according to equivalence. The authors conclude that these data show that "class-based behavior generalized between two trial formats; matching-to-sample trials during class formation, and sorting during post-class formation testing" (p. 30).

In other studies, more discordant results were found (e.g., Arntzen et al., 2017; Eilifsen & Arntzen, 2009; Fienup & Dixon, 2006; Green, 1990; Pilgrim & Galizio, 1996; Smeets et al., 2000). Eilifsen and Arntzen (2009) studied the trial types role in tests for stimulus equivalence. The stimuli consisted of nine arbitrary, abstract and black symbols organized into three classes with three members. The participants were given all the pictures and asked to categorize them in the pre-class formation sorting test, and after the experiment, they were again told to categorize the stimuli. The results of the study, with respect to the sorting tests, were that none of the participants placed the stimuli into categories like the experimenter-defined classes in the pre-sorting, showing that they did not know what the classes were prior to the experiment. In post-categorization, 14 of the 20 participants sorted the stimuli corresponding to the experimenter-defined classes. Six of these 14 participants did not respond according to stimulus equivalence in the prior MTS test, and one participant who responded according to stimulus equivalence in the MTS test did not sort the stimuli according to the experimenter-defined classes. The authors of the article concluded that the "stimulus sorting did not converge well with responding to the test for stimulus equivalence" (p.199) and explained that this could be due to the differences between the two test types. The sorting test was done in a single trial. All stimuli were present simultaneously, and the participant could "scan back and forth between the different stimuli" (Eilifsen & Arntzen, 2009, p. 199).

Maintenance of Stimulus Classes

In several of the articles the pre-class formation sorting test showed the maintenance of equivalence classes (e.g.,Arntzen, 2004; Arntzen et al., 2017; Arntzen et al., 2014; Arntzen et al., 2015; Cowley et al., 1992; Fields et al., 2014; Fields et al., 2012; Hove, 2003; Nartey et al., 2014, 2015; Nedelcu et al., 2015; Smeets & Barnes-Holmes, 2005). In Eilifsen and Arntzen (2009) it was found that one participant did not sort the stimuli into the experimenter defined classes, having reached the criteria in the prior equivalence test, thereby not showing maintenance of stimulus classes.

Delayed Emergence

We have seen in several of the previously described examples that although the post-class formation sorting test maintained the classes formed in the previous test in MTS format testing for the experimenter-defined equivalence classes, the post-class formation sorting test produced more intact experimenter-defined classes than the test in the MTS format did (e.g., Arntzen et al., 2017; Arntzen et al., 2015; Eilifsen & Arntzen, 2009; Fields et al., 2014; Fields et al., 2012; Fienup & Dixon, 2006; Hove, 2003; Smeets & Barnes-Holmes, 2005; Smeets et al., 2000).

Research on Sorting

In the study by Fields et al. (2014) on stimulus sorting, fifty students attempted to form three 5-member classes using an MTS-based training procedure in an LS training structure; afterwards they were tested in an MTS-based procedure. Subsequently, the participants were given a sorting test with cards. The participant got the instruction to put the cards into groups. Twenty-four of the participants showed equivalence classes of the experimenter-defined classes under the MTS-based test. Of the 24 participants, 23 also demonstrated maintenance of the experimenter-defined classes under the subsequent sorting performance. In addition, none of the participants who did not form the equivalence classes in the test for emergent relations sorted the stimuli according to the experimenter-defined classes. It was concluded that “on a group and a within-subject basis, sorting tests appear to provide a valid measure of the maintenance of equivalence classes, or lack of class formation” (p.494).

A study by Arntzen et al. (2015) was conducted to determine whether a sorting test could be used to measure the immediate emergence of equivalence classes. Sixteen participants were trained in baseline relations of three 5-member classes using an LS and MTS format. Afterwards, the participants were tested with two MTS tests, one after the other. Three participants showed immediately emergence of the equivalence classes, and two participants showed the emergence of classes in the second MTS test. However, two other participants had a long delay in the emergence of classes, and this was first shown in the following sorting test. The three participants with an immediate emergence of equivalence classes were trained in another set of stimuli, then immediately tested with the sorting test and afterward the two MTS tests. These participants showed immediate emergence of the equivalence classes by sorting the stimuli according to the experimenter-defined classes. In a sorting test, all emergent relations in an equivalent class are not tested. The design of the study included evaluation of whether the formation of the stimulus classes in the sorting test could be considered equivalence classes, by conducting the two MTS tests containing all the derived relations of baseline, symmetry, transitivity and equivalence from the classes. The authors concluded, “this experiment represents the first demonstration of the use of a sorting test to document equivalence formation” (Arntzen et al., 2015, p. 624). Additionally, in the first part of the experiment, a delayed emergence of the classes was measured by the sorting test, and the authors concluded, “it is possible that the sorting test was more sensitive to the delayed emergence of the classes than the traditional MTS test” (Arntzen et al., 2015, P. 624).

In the study of Arntzen et al. (2017), the relation between the MTS tests and sorting tests were investigated with regard to the formation of equivalence classes. A design was created that could test the sensitivity of the sorting test as a measure for class formation and equivalence classes. Twenty college students participated and were randomly assigned to two groups. Both groups trained the baseline relations of three classes with five members in a linear series format, all with abstract stimuli, and all were presented on the computer screen. After the training Group 1were exposed to the sorting test first, then the MTS test, and finally a second sorting test. Group 2 were exposed to the MTS test first, then a sorting test, and finally a second MTS test. It was found that in Group 1, 50% of the participants showed immediate emergence of the three classes with the sorting test, and only 30% reached the criteria in the subsequently MTS test for emergent relations. In Group 2, 30% of the participants showed immediate emergence with the MTS test, and in the subsequent sorting test, as many as 60% of the participants documented all the experimenter-defined classes. The authors of the article concluded that the findings were not completely concordant, and they suggested that this could be due to procedural variables.

Economy of Sorting Tests

In several studies, the effectiveness of the sorting test is described. For example, it is found that the sorting test was completed in less than 5 minutes on average, whereas the MTS test only required approximately 25 -30 minutes to administer (e.g., Arntzen et al., 2015; Fields et al., 2014; Fields et al., 2012).

Discussion

Both concordant and discordant findings on the sorting tests ability to measure the MTS test corresponding equivalence classes were found in this review.

In considering the use of sorting tests as post-class formation tests for an additional measurement of class formation, we observed in the early studies of equivalence that there is little doubt that the sorting test can show the maintenance of stimuli classes. We have also seen in the review that the sorting test often showed what has been called a delayed emergence of stimuli classes. Additionally, in some of the studies where the experimenter-defined classes were first shown in the post class formation sorting test after the MTS-based test, the question was raised: if the sorting test is a more sensitive test to the delayed emergence of stimuli classes.

Some of the latest research has shown that sorting can measure the immediate emergence of equivalence classes. Additionally, a sorting test that is in accordance with experimenter-defined classes definitely documents the emergence of arbitrary stimulus classes. However, the sorting test does not necessarily document the derived relations of symmetry, transitivity and equivalence. To clarify whether the sorting tests documented emergent relations, an MTS test was conducted after the sorting test by Arntzen et al. (2015). They found in three out of three participants that sorting tests could most likely document the immediate emergence of equivalence classes. On the other hand, research also revealed — through the MTS-based test for emergent relations after the sorting test — that the classes could not be defined as equivalence classes for all participants (Arntzen et al., 2017). For other participants, the MTS-based tests after the sorting tests showed equivalence classes, and these classes could well have had the same functions in the sorting test as in the MTS test (Arntzen et al., 2017). The discordant results of the research opened up the additional question about the function of the classes documented by MTS tests and sorting tests; if they can be assumed to have the same functions. To address this question further, Arntzen et al. (2015) proposed investigating the functions of stimulus classes documented by sorting tests to reveal whether the classes have some of the properties of the equivalence classes. For example, when a stimulus class is documented by sorting to investigate whether a new response trained to one member of the stimulus class, would generalize to the remainder of members of the stimulus class, as is shown for equivalence classes in other experiments (e.g., Arntzen, Eilertsen, & Fagerstrøm, 2016; Augustson & Dougher, 1997; Augustson, Dougher, & Markham, 2000; Fields & Garruto, 2009). Additionally, Arntzen et al. (2017) proposed to investigating whether a higher correspondence between the sorting yields and MTS outcomes could be found by extending the baseline training.

Differences in Variables and Parameters

We found in the review that the participants assigned to the different studies varied in age from two to 62 years and from having mental disability to university students. Arntzen (2012b) stated about the participants age that it seems to be a difference in effects on different training structures in MTS. If age yields a difference in results on sorting tests we do not know. But, in the latest research on sorting (Arntzen et al., 2017; Arntzen et al., 2015; Fields et al., 2014) recruited adult participants, mostly college students. Thus, it is not plausible that the age of the participants can describe the discordant findings between those studies.

It is reasonable to surmise that the differences in the procedural variables that we found in sorting procedures — for example, with regard to the differences in the type of stimuli used, the presentation of the stimuli, the differences in the instructions used before sorting and how the participants are supposed to organize the stimuli during the sorting test — can influence the yields on the sorting tests. Regarding the different types of stimuli used (see Table 4), it was used abstract stimuli of the same type in the latest studies on sorting (Arntzen et al., 2017; Arntzen et al., 2015; Fields et al., 2014) thereby discordant findings in those studies is not to be found in the variation of stimuli. In other studies stimuli with appearances such as a spoken word, a written name and a picture are undoubtedly arbitrary and non-identical, but it could theoretically be argued that stimuli like Greek letters or other abstract shapes drawn in black on a white background have some features that is close to identical. It could be argued that it is always possible to find identical parts in the lines and shapes, even if they are tiny parts. If participants look for or observe identical parts in the stimuli when forming stimulus classes and the possible influence on the outcome of the MTS test and the yields of the sorting test, it would be possible to test empirically by using compound stimuli with identical parts.

Regarding how participants are supposed to organize the stimuli during the sorting test, we need to consider that in a sorting test conducted with a randomly shuffled deck of cards, the probes that are evaluated will be randomly chosen from participant to participant. Additionally, in a sorting test, fewer probes are involved than in an MTS-based test for emergent relations. An MTS-based test for emergent relations assess and give a defined number of probes for all the emergent relations of symmetry, transitivity and equivalence relations. Sidman (1994) described that a class of N stimuli contains N^2 relations of two stimuli. Of these is $N^2 - 2N - 1$ emergent relations (Fields & Verhave, 1987). However, in a class of N stimuli, a sorting test can document N - 1 relations of the relations of the equivalence class. As we have seen, the sorting tests have been conducted in different ways. In some studies, the participants were told to put the cards in stacks, and in others, they were told to form clusters or arrays. When forming clusters, the participants are able to scan all cards simultaneously, whereas when the participant is sorting the cards into stacks, and if the stacks are locked after a card is placed, the participant would initially only be able to see the upper card in the deck of cards, after which the participant would place this upper card on the table as the top card of the stimuli from a class. Thereafter, with the next card from the deck of cards, the participant would either match the card with the top card of a stack already placed on the table or make a new stack; this would continue until the deck of cards is empty. Thus, sorting each of the cards into stacks is controlled by only one stimulus from each class. Theoretically, the last procedure provides a stronger assessment of class formation (Arntzen et al., 2017). In the reviewed articles stating that the sorting was done in stacks, none specified whether the stacks were locked. However, using the locked stack procedure with cards or other tangible items, it would be more challenging to lock the stacks than if the same procedure is perform on a computer screen. Therefore, performing a sorting task on a computer screen with locked stacks would theoretically provide an even stronger assessment of class formation. This theory will have to be tested empirically to reveal its validity.

The numbers of classes used in the studies varied from two to six and the numbers of members in the classes varied from three to seven. The most common number of classes and members was three 3-member classes used by 44% of the studies. About the effects on MTS tests Arntzen (2012b) stated that “it is not clear how an increasing number of members and/or number of classes influences the emergence of equivalence classes depending on training structure” (p.125). How different numbers of classes and numbers of members influence the soring test is not clear either. But, keeping this variable steady by using the same numbers of classes and members between the studies is a possibility to control for potential different effects. The latest studies on sorting all used three 5-member classes.

We found that many different instructions were given before the sorting test (see Table 7). Instructions on how to conduct the sorting task appears to be a necessity, but just as (Sidman, 1992) issued a warning about instructions used for MTS procedures — because we do not know whether it is rules “that give rise to equivalence, or equivalence that makes rules possible” — the same reasoning can be used for the sorting procedures. If we want to ensure that we are measuring the experimental effects and not the participant’s verbal history, we should minimize the use of instructions. One possible way to conduct an experiment that involves minimal verbal instructions for both the sorting and the MTS procedure could be to show the participant a video of similar procedures but with other stimuli. For the sorting test, the video should be filmed from an angle that does not show the whole computer screen — so the participant cannot figure out how many stacks of cards are on the screen — and play parts of the video quickly so the participant cannot figure out how many cards are moved from the deck of cards. Whether this procedure would affect the outcomes for the MTS tests and yields for the sorting tests can be tested empirically.

We found that it was used the linear series training structure in 14 out of 25 articles. The different training structures (MTO vs. OTM) were shown by Hove (2003) to give different outcomes for both the outcome of the MTS-based test of emergent relations and the yields of the post-class formation sorting test. Hove found that MTO yielded a higher outcome in the MTS-based test for emergent relations compared to OTM, and MTO also had the highest yields in the post-sorting test. The same results for MTO as the training structure before the MTS-based test was found by Saunders, Chaney, and Marquis (2005), whereas

Arntzen, Grondahl, and Eilifsen (2010), did not find that MTO was superior to OTM. Therefore, more research is needed on the training structures of MTO and OTM to clarify the issue of the outcome on MTS tests and sorting tests.

Other differences in the MTS training procedures were seen (see Table 1) for example in the training protocol. Different procedures in the training protocol is known to influence the outcomes on MTS-based tests (Arntzen, 2012b). Additionally, in the arrangement of the training trials, number of nodes, and test trials, we observed differences. These differences in procedures might influence the yields of the post-class formation sorting tests the as well as the outcome of the MTS test. These variables were not further reviewed. To uncover and determine whether any of these differences in the parameters are responsible for the discordant results between the MTS tests and the sorting test, an extended review should be completed.

We found in the three studies on sorting that it was used some of the same variables and parameters for example considered participants, stimuli, training structure and number of classes and members. However, there was a difference in the conduction of stimuli sorting and in the arrangement of training trials in the prior MTS training. Two studies used a tabletop procedure with cards and one a computer screen procedure and two studies used a concurrently presentation of training trail one a serialized. If this differences in procedure was the reason for discordant findings between the studies is unknown, and we suggest further investigation to reveal if and how the different variables and parameters in both the prior MTS training and in sorting tests influence the results on the sorting tests.

Limitations

A limitation in the present review is that we only conducted searches in PsycINFO; it is possible that supplementary searches in other databases would have yielded more articles.

Further Research

At present it seems most important, to clarify the differences in the functions between the classes documented with MTS-based tests for emergent relations and those documented with sorting tests. One function that could be investigated is whether a new response trained to one member of a stimulus class that is documented by a sorting test would be generalized to the remainder of the members of that stimulus class, as is shown for equivalence classes.

It also appears important to clarify through research whether it is possible to achieve a higher correspondence between the two test types by changes in procedures — for example, by extending the baseline training or by using stacking with locked stacks as a computer screen procedure instead of a grouping procedure on a table.

Conclusion

Being able to measure class formation and the emergent relations with the sorting test only would clearly be beneficial in relation to time saving, compared to the more time consuming MTS-based test on emergent relations. In the review, we have studied different procedures used when conducting a sorting test. We have focused on some of the variables and parameters used in the sorting procedures and in the MTS-based training procedure.

We have seen that immediate emergence of equivalence classes has been documented in the sorting tests in two experiments and some studies support the understanding that equivalence class formation can be documented with sorting tests. We found that sorting tests appear to track more classes than the MTS-based equivalence test and that this was described as either delayed emergence or the sorting test being a more sensitive measurement of class formation. In some of the latest research papers, we also found that not all classes documented by sorting tests can be defined as being functionally the same as the classes documented by the MTS tests for emergent relations. To empirically reveal how these functionally different classes differ is essential.

References

Arntzen, E. (2004). Probability of equivalence formation: Famliar stimuli and training sequence. *The Psychological Record, 54*(2), 275–291. Retrieved from http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1235&context=tpr.

Arntzen, E. (2012a). Om stimulusekvivalens: Teoretiske betraktninger, oppsummering av en del empiri og noen praktiske implikasjoner. In S. Eikeseth & F. Svartdal (Eds.), *Anvendt atferdsanalyse: Teori og praksis* (2 ed., pp. 100–138). Oslo: Gyldendal akademisk.

Arntzen, E. (2012b). Training and testing parameters in formation of stimulus equivalence: methodological issues. *European Journal of Behavior Analysis, 13*, 123–135. Retrieved from http://www.ejoba.org/.

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. doi:10.1080/15021149.2011.11434398

Arntzen, E., Eilertsen, J. M., & Fagerstrøm, A. (2016). Preferences in equivalence classes by low potency benign valenced stimuli. *European Journal of Behavior Analysis, 17*(2), 142–153. doi:10.1080/15021149.2016.1247637

Arntzen, E., Granmo, S., & Fields, L. (2017). The relation between sorting tests and matching-to-sample tests in the formation of equivalence classes. *The Psychological Record*. doi:10.1007/s40732-016-0209-9

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://thepsychologicalrecord.siuc.edu/index.html.

Arntzen, E., Nartey, R. K., & Fields, L. (2014). Identity and delay functions of meaningful stimuli: Enhanched equivalence class formation. *The Psychological Record*, 349–360. doi:10.1007/s40732-014-0066-3

Arntzen, E., Norbom, A., & Fields, L. (2015). Sorting: An alternative measure of class formation? *The Psychological Record, 65*, 615–625. doi:10.1007/s40732-015-0132-5

Augustson, E. M., & Dougher, M. J. (1997). The transfer of avoidance evoking functions through stimulus equivalence classes. *Journal of Behavior Therapy and Experimental Psychiatry, 28*, 181–191. Retrieved from ISI:A1997XZ11200002.

Augustson, E. M., Dougher, M. J., & Markham, M. R. (2000). Emergence of conditional stimulus relations and transfer of respondent eliciting functions among compound stimuli. *The Psychological Record, 50*, 745–770. Retrieved from ISI:000165227100008.

Cowley, B. J., Green, G., & Braunling-McMorrow, D. (1992). Using stimulus equivalence procedures to teach name-face matching to adults with brain injuries. *Journal of Applied Behavior Analysis, 25*, 461–475.

Dickins, D. W. (2011). Transitive inference in stimulus equivalence and serial learning. *European Journal of Behavior Analysis, 12*, 523–555. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=510.521.521.665.2613&rep=rep2611&type=pdf.

Dickins, D. W. (2015). A simpler route to stimulus equivalence? A replication and further exploraiton of a " simple discrimination training procedure" (Canovas, Derbert and Pilgrim 2014). *The Psychological Record, 65*(4), 637–647. doi:10.1007/s40732-015-0134-3

Dymond, E., & Rehfeldt, R. A. (2001). Supplemental measures and derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin, 19*, 8–12. Retrieved from http://www.eahb.org/11/wp-content/uploads/2014/2003/Dymond-Rehfeldt.pdf.

Eilifsen, C., & Arntzen, E. (2009). On the role of trial types in tests for sitmulus equivalence. *European Journal of Behavior Analysis, 10*, 187–202. Retrived from http://www.ejoba.org.

Eilifsen, C., & Arntzen, E. (2011). Single-subject withdrawal designs in delayed matching-to-sample procedures. *European Journal of Behavior Analysis, 12*, 152–172. Retrieved from http://www.tandfonline.com/doi/abs/110.1080/15021149.15022011.11434361.

Fields, L., Arntzen, E., & Moksness, M. (2014). Stimulus sorting: A quick and sensitive index of equivalence class formation. *The Psychological Record, 64*, 487–498. doi:10.1007/s40732-014-0034-y

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

Fields, L., & Garruto, M. (2009). Optimizing linked perceptual class formation and transfer of function. *Jounal of the Experimental Analysis of Behavior, 91*, 225–251. doi:10.1901/jeab.2009.91-225.

Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimantal Analysis of Behavior, 48*, 317–332. doi:10.1901/jeab.1987.48-317

Fienup, D. M., & Dixon, M. R. (2006). Acquisition and maintenance of visual-visual and visual-olfactory equivalence classes. *European Journal of Behavior Analysis, 6*, 87–98. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.81.81.664.8771&rep=rep8771&type=pdf.

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

Hove, O. (2003). Diferential probability of equivalence class formation following a one-to-many versus a many-to-one training structure. *The Psychological Record, 53*, 617–634. Retrieved from http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1501&context=tpr.

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

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. doi:10.1901/jeab.2002.78-527

Ludvigson, H. W., & Caul, W. F. (1964). Relative effect of overlearning on reversal and nonreversal shifts with two and four sorting categories. *Journal of Experimental Psychology, 68*, 301–306. doi:10.1037/h0042254

Mackay, H. A., Wilkinson, K. M., Farrell, C., & Serna, R. W. (2011). Evaluating merger and intersection of equivalence classes with one member in common. *Journal of Experimental Analysis of Behavior, 96*, 87–105. doi:10.1901/jeab.2011.96-87.

Nartey, R. K., Arntzen, E., & Fields, L. (2014). Two discriminative functions of meaningful stumili that enhance equivalence class formation. *The Psychological Record, 64*, 777–789. doi:10.1007/s40732-014-0072-5

Nartey, R. K., Arntzen, E., & Fields, L. (2015). Enhancement of equivalence class formation by pretraining discriminative functions. *Learning & Behavior, 43*, 20–31. doi:10.3758/s13420-014-0158-6

Nedelcu, R. I., Fields, L., & Arntzen, E. (2015). Arbitrary conditional discriminative functions of meaningful stiuli and enhanced equivalence class formation. *Journal of the Experimantal Analysis of Behavior, 103*, 349–360. doi:10.1002/jeab.141

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

Rosch, E., & Mervis, C. (1977). Children´s sorting: A reinterpretation based on the nature of abstraction in natuarl categories. In R. C. Smart & M. S. Smart (Eds.), *Readings in child development and relationships* (2 ed., pp. 140–148). New York: MacMillan.

Saunders, R. R., Chaney, L., & Marquis, J. G. (2005). Equivalence class establishment with two-, three-, and four-choice matching to sample by senior citizen. *The Psychological Record, 55*(4), 539–555. Retrieved from https://login.ezproxy.hioa.no/login?URL=?url=http://search.proquest.com/docview/212766566?accountid=212726439.

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

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

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

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

Sigurdarddottir, 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 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3363409/.

Smeets, P. M., & Barnes-Holmes, D. (2005). Auditory-visual and visual-visual equivalence relations in children. *The Psychological Record, 55*(3), 483–503. Retrieved from http://search.proquest.com/docview/212764319?pq-origsite=gscholar.

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/cgi/viewcontent.cgi?article=1401&context=tpr.

Travis, R. W., Fields, L., & Arntzen, E. (2014). Discriminative functions and over-training as class-enhancing determinants of meaningful stimuli. *Journal of Experimental Analysis of Behavior*. doi:10.1002/jeab.91

Varelas, A., & Fields, L. (2017). Equivalence based instruction by group based clicker training and sorting tests. *The Psychological Record, 67*, 71–80. doi:10.1007/s40732-016-0208-x

Table 1

Matrix Showing Parameters and Variables in Reviewed Articles

	Authors	Year	Journal	No.of participants/ Population details/ Age in years	Numbers of classes and members Stimuli used	Some parameters from the procedures of sorting	Some parameters from the procedures of training baseline relations	Some parameters from the procedures of testing equivalence	Quotes and comments about findings and conclusions on sorting
1	Arntzen	2004	TPR	50 / College students / ?	Three classes of five members. Group 1: All Greek and Arabic letters Group 2: Same but pictures as A-stimuli Group 3. Same but pictures as E-stimuli Group 4: Nonsense syllables as A-stimuli Group 5: As Group 1(but a response was required - a touch on the keyboard)	Sorting with printouts of the stimuli. Categorization in the end of the experiment. Participants were asked to categorize the stimuli.	Training in MTS format on computer screen. Many-to-one training. Baseline relations established successivly. Trials of simultaneous matching. First trials with only the correct comparison gradually increasing to three comparisons. The Sample on the left hand of the scren. The comparisons on the right hand of the screen. A response to sample needed for apperance of comparisons stimuli. Programmed consequences. 24 correct trials before testing.	Testing in MTS format on computer screen. Two blocks of testing. Four of each trial type in each test block	"Participants who responded in accord with equivalence also categorized the stimuli in the three experimenter-defined groups, while 23 participants who did not respond in accord with equivalence did not categorize the stimuli in the three experimenter defined groups". (p. 284)
2	* Arntzen, Braaten, Lian, Eilifsen	2011	EJOBA	20/ adults/ 22–39	Two stimulus set, both: Four classes with three members. All abstract stimuli. Black on White background.	Sorting conducted with cards. Sorting before and after the experiment. Were asked to categorize them.	Training conducted on a computer. Linear series training in MTS format. Simultaneous protocol. Concurrently discrimination. Training trials presented randomly. Each trial type presented five times. Blocks consisting of 40 trials. Criterion to proceed to next phase: 90%. Training phase repeated until criterion reached. Programmed consequences. Maintenance of baseline trials:Block with 75%, 25% and 0% programmed conseqences.	Testing conducted on a computer A total of 120 trials. No programmed consequences.	" The data shows that the participants did correct categorization of the stimuli cards even if they did not show formation of equivalence classes during the testing ." (p.519)
3	* Arntzen, Granmo, and Fields	2017	TPR	20 / College students / 20-44	Three classes of five members. All abstract stimuli. Black on white background.	Sortingn on computer screen. All 15 stimuli randomized in one stack. Instructions: "Drag the top stimulus to the side" and told to "drag all stimuli apart form each other so that all stimuli were visible on the screen in front of you" and to " put them togheter in the way that you think is correct, but make sure that all stimuli are visible"."	Training in MTS format on computer screen. Linear series training. Baseline relations established concurrently. Trials of simultaneous matching. The sample in the middle of the screen. The three comparisons in the corners of the screen. A response to sample needed for apperance of comparisons stimuli. Programmed consequences. Blocks of 60 trials. Blocks repeated until all trials correct. Training for maintenance: Block with 50% and then block with no programmed consequences. 100% mastery criterion.	Testing in MTS format on computer screen. Two mixed test blocks. Each block contained 180 trials (36 baseline, 36 symmetrical, 54 transitive and 54 equivalence relations). Randomized order. No programmed consequences. Mastery criterion 95% of all the trials in each block.	"Sorting tests can be used to track the immediate emergence of all the stimulus classes included in the design.Specically, this occurred with eight participants." "The results of this experiment produced a substantial increase in the prevalence of the immerdiate emergence of stimulus classes using a sorting test." (p.94)

4	* Arntzen, Nartey & Fields	2014	TPR	40 / University students / 19–31	Three classes of five members Three conditions: In Two conditions, all abstract stimuli black on white background. In one condition, C stimuli as meaningfull .	Sorting conducted with plastic laminated cards. Participants sorting the stimuli used in the condition they were assigned to. Told to "put them into groups".	Training in MTS formate on computer. Linear series training structure. Presented concurrently (every trial type trained from the beginning of a block). Ramdomly presented. 36 trials per block. Mastery criterion 90% correct. If participant did not meet criterion, the block repeated until criterion met. Maintenance of baseline trials. Block with 75%, 25% and 0% programmed conseqences.	Testing conducted on a computer One test block of 180 trials. (36 baseline, 54 one-node trials, 36 two-node trials, 18 three-node trials) No programmed consequences Mastery criterion defined as 90% of the trials in the block.	"Pre-class formation sorting performance".."no participants sorted the stimulus cards ito piles that corresponded to the three experimenter-defined classes." and "...,the data obtained form 39 of the 40 participants in the experiment showed perfect concordance between the performances evoked by the derived relations and the sorting tests." (p.355)
5	Arntzen, Norbom, and Fields	2015	TPR	16 / Individuals / 20-43	Two sets of three classes of five members. All abstract stimuli. Black on white background.	Sorting with a deck of cards presented on the computer. Instructions: "Put these into groups as you felle like." (p.619) Participants were informed to move the stimuli on the screen to form clusters. Afterwards the experimenter took a screenshot of all the stimuli on the screen.	Training in MTS format on a computer screen. Linear series training. Baseline relations established concurrently. Trials of simultaneous matching. The sample in the middle of the screen. The three comparisons in the corners of the screen. A response to sample needed for apperance of comparisons stimuli. Programmed consequences. Blocks of 60 trials. Blocks repeated until responding correctly on at least 90% of trials in block. Training for maintenance: Block with 75% consequences readministered to 90% correct and then block with 25% programmed consequences readministered to 90% correct at last no programmed consequences.	Testing in MTS format on computer screen. Two test blocks of 300 test trials (60 baseline trials, 60 symmetry trials, 90 transitivity trials and 90 equivalence trials). Randomized order. No programmed consequences. Mastery criterion 95% of all the trials in each block.	" This experiment represents the first demonstration of the use of sorting test to document equivalence class formation. This finding reduces any procedural constraints that have to be considered when defining equivalence classes. Sorting tests were completed 10 times faster than a typical MTS-based test for class formation. Thus, the use of sorting tests should significantly reduce the time needed to establish equivalence classes in applied and basic research settings. The sorting test tracked class formation by sampling a subset of emergnt relations. Thus, documenting all emergnet relations may not be needed to show equivalence class formation." (p.625)
6	Cowley, Green, and Braunling. McMorrow	1992	JABA	3 / Men with brain injuries / 45, 30, 57	One set of stimuli. Three classes with three members: A :Dictated names B: Photographs of faces C: Written names D: Nameplate photograph (for one participant)	Pretraining and testing in a table-top match-to-sample format. Sorting with the stimuli handed to the participants. Pre and post testing Instructions: "Put all the cards that go togheter in stacks". (p. 466)	The table-top match-to-sample only the AB relation. Sample were visual stimuli or dictated words. Comparisons were always three visual stimuli Praise as consequence. 30 trials in each set of matching pretest trials. Pretraining in match-to sample formate - without reinforcement. Pretraining with oral naming and sorting to.	In a match-to sample format. Without differential consequences. Tested relations: BC/CB or BC/CB and BD/DB(for one participnat)	One participant´s "score on the sorting pretest was 81%, but neither of the other participants sorted stimuli accuratly."(p.469) "All participants demonstrated the formation of three equivalence classes, each including a therapists dictated name, photo and written name." (p.471) One participant was discharged from the institution before completing sorting posttests. (p.471) Two participants grouped all stimuli correctly in the sorting posttest.

7	Dickins	2011	EJOBA	20 / Psychology undergraduates / 18-35	Two classes with seven members. Members having the transitive inference: If A>B and B>C, then A>C and so on. Passport photographs of privious students. 7pictures of students from the discipline of Psychology and 7 pictures from the disipline of Zoology. The participants told they were presented with pairs of students from the same discipline with the one to the left having one higher scoore than the one to the right on the final exam.	Sorting conducted on a computer screen. Three sorting phases - used for testing categorization of both classes and order of stimuli. 1. Sorting test right after training. 2. Sorting test after equivalence or serial learning tests. 3. Sorting test after swapping test types between the groups. Sorting by draging the stimuli with the mouse. The participants were asked to arrange the photoes into two groups in a matrix of 35 squares.	Training conducted on computer screen. 1. Introduction to be read. 2. A study phase: of the 12 pairs to watch and learn. The pairs were presented successively but randomly. Alternated with: 3. Response phase: different in two groups. Group 1. Stimulus equivalence group: Matching-to-sample trials. Group 2. Serial learning group: Study pairs with indication. No feedback, but at the end of the response phase. If one or more incorrect another study phase were given, until learning criterion of 12 /12 obtained.	Testing conducted on computer screen. Unreinforced. Equivalence test trails or serial learning test trials. 36 trials for each nodal number. 36 trials of 'zero-nodal' premise. SE Group: First tested in MTS format afterwards in SL format SL Group: First testet in SL format afterwards in MTS format.	"Results on the sorting task gave stronger indications of set formation than the transitive inference tests, particulary in the SL first group" (p. 523)
8	Dickins	2015	TPR	Experiment 2: 4 / Psychology undergraduate / 18–19 Experiment 3: 5 / Psychology undergraduate / 18–20	Experiment 2 and 3: Phonologically correct non- words as printed sample stimuli and as spoken responses. In second experiment: two classes of 5 members. In Third experiment: Three classes of 4 members.	Conducted on a computer. Conducted in the second and third experiment. Conducted on an empty matrix of 35 squares, surrounded by the ten non-words used, pluss eight words not used in the experiment. Instructed to "drag the words they remembered had been used into separate cells in the table, putting them into some kind of order" (p.6)	Training conducted on a computer. Three experiments. In all experiments: First discrimination training In blocks of 12 trials. Blocks presented until 100% correct in two successive blocks. Training new responses (new key pressing) to two of the stimuli.	First Transfer-of-function test. Then Compound stimuli test in a Go/ No-go procedure. Two-choise MTS test.	Participant C11: "thoug he only imperfectly recalled four or five of the other stimulus words, he grouped the stimuli perfectly in the sorting test" Participant C12: "was able to recall all the stimuli and their interrelations verbally, except for one stimulus word, and performed perfetly on the sorting test". Participant C13: "was only able to recall five of the stimui verbally. Her sort was perfect, however". Participant C14: "she promptly recalled 11/12 words and sorted them perfectly. Particpant C15: "recalled 11/12 of the words in correct groupings and performed perfectly on the sorting test". (p.8)
9	Eilifsen, and Arntzen	2009	EJOBA	20 /Adult people / 19-42	Three classes with three members. Nine abstract symbols in black on white background.	Sorting conducted with the same stimuli as plastic laminated pictures. To controle for categorization before and after experiment. Before participants began the experiment on the computer they were given plastic laminated pictures of the stimuli and told to categorize them. After experiment participant were again given the pictures to categorize them.	Training conducted on a computer. MTS format with linear series training structure. Baselines relations established concurrently. Trials of simultaneous matching. The sample in the middle of the screen. Three comparisons in the corners of the screen. A response to sample needed for apperance of comparisons stimuli. Programmed consequences. Blocks of 18 trials. Blocks repeated until 16 of 18 correct. Then when participants reached criterion feedback was reduced to 75% then 50% then 25% and afterwards 0%. The phase was presented again when preformance was below criterion.	Testing conducted on a computer. One mixed testblock. 54 trials. Randomized order. Testing all relations of directly trained, symmetry, transitivity and global equivalence three times. No programmed consequences. Mastery criterion 17 out of 18 of all types of relations. If not reaching the criteria the participants were exposed to training again.	Pre-sorting: No participants sorted in accordance to experimenter defined classes. Post-sorting: 14 participants sorted in accordance to experimenter defined classes. 6 of these did not reach criterion for equivalence in the prior MTS test. One participant who reached the criteria in the equivalence test did not sort the stimuli in accordance to experimenter defined classes. "Thus, stimulus sorting did not converge well with responing on the test for stimulus equivalnece" (p.199)

10	Eilifsen, and Arntzen	2011	EJOBA	10/ Adult people / 20-28	Three sets of stimuli: Three classes with five members. Abstract symbols in black on white background.	Sorting conducted with the same stimuli as printed copies. Pre-class formation sorting and post-class formation sorting. Participants were asked to categorize the stimuli as they liked. The three sets categorized separately. Conducted with plastic laminated cards.	Training conducted on a computer. Two conditions on the startingpoint for titrating the delayed MTS. MTS format with linear series training structure. Simultaneous protocol. Concurrent introduction of the stimulus relations. The sample in the middle of the screen. The comparisons in the corners of the screen. A response to sample needed for apperance of comparison stimuli. Programmed consepuences. Blocks of minimum 144-36-36-36-36 trials.	Training conducted on a computer. Mixed testblocks after each condition. 180 trials. Randomized order. Testing all relations of directly trained, symmetry, transitivity and global equivalence three times. No programmed concequences. Mastery criteria 90% for each relation.	"In the soritng task of the start of the experiment none of the participants categorized the stimuli according to any of the experimenter defined stimulus classes". (p. 165) No other conclusion on sorting.
11	* Field, Arntzen, and Moksness	2014	TPR	50 / College students/ 19-62	Three classes of five members. Abstract symbols in black on white background.	Sorting was conducted with the same stimuli as printed copies. Pre-class formation sorting and post-class formation sorting. Participants were instructed: "please, put them into groups and call me when you have completed the task". (p.489)	Training conducted on a computer. Linear series training structure in MTS format. Trials of simultaneous MTS format. Baseline relations trained in a serial order. The sample i the miiddle of the screen. The comparisons in the corneres of the screen. A response to sample needed for apperance of comparison stimuli. Programmed consequences. Training a relation for all classes until all trials in blocks mastered 100%.	Testing conducted on a computer. Test block with all baseline, symmetry, transitivity and equivalence relations. Test block repeated up to two times or to at least 90% of trials and relations in each class met mastery criterion of 90%.	43 of 50 participants showed one-to-one correspondence between the presence/absence of the three classes during the MTS-based derived-relations and sorting tests. "these results support the view that sorting data provide a valid measure of the mainteenance of equivalence classes" (p.493) "In some cases there were indications that the sorting was more sensitive to class formation than MTS-based derived-relations test" (p.495)
12	Fields, Arntzen, Nartey, and Eilifsen	2012	JEAB	30 / College students / 19-45	Three sets of stimuli: Three classes with five members en each set. Nine extra stimuli used for discrimination training of the abstract C stimuli. Abstract symbols in black on white background in two groups and C stimuli as meaningful stimuli in one group.	Sorting was conducted with the same stimuli as printed copies. Pre-class formation sorting task and post-class formation sorting test. Iinstructions were: "Please put them into groups and call me when you have completed the task". (p.168)	Training conducted on a computer. Three groups each with different stimuli. All groups underwent: MTS format with linear series training strucctur. Five serializes phases. The sample in the middle of the screen and three comparisons in the corners of the scren. A response to sample needed for apperance of comparisons stimuli. Programmed consequences. Blocks of 36 trials. Block repeated until correct response on at least 90% of trials of each type. Numbers of training trails equalized. Maintenance of baseline relations: The percentage of trials with feedbask was redused to 75%, 50%, 25% and 0 % before testing. One group underwent discrimination trainng of C stimuli before testing for emergent relations.	Testing conducted on a computer Two test blocks of 90 trials (18 baseline trials, 18 symmetry trials, 27 one-node trials and 18 two-node trials and 9 three-node trials). Presented randomly. No feedback Mastery criteria defined as 90% of trials for each type of relation.	"The participants who formed classes, as indexed by the derived relations test performances, also showed the intactness of the same classes on the postclass formation card sorting tests. Thus there was a perfect association of the performances produced by the two type of test." (p. 171) "Equivalence classes was documented by performances seen in the card sorting-tests but not in the emergent relation tests" (p.173)

13	Fienup and Dixon	2006	EJOBA	4 / College students/ 21-25	Two sets of stimuli. Each set of stimuli included three classes of stimuli with 3 members. All visual stimuli were printed on 5x7 cm cards and laminated. The the stimuli were pictures of patterns found on sweaters, chairs, and rugs. Pictures of objects like rocks, and cement were used as B and C stimuli. Set one contained olfactory stimuli as A stimuli. Stored in similar plastic bottles.	Sorting as the last phase to test for stimulus class membership. Instructions: "place the objects in three piles" (p.92)	Half of group, start training with set 1 other half set 2, they swap later. OTM training. Praise as feedback. Blocks of nine trials. Criterion for mastery: Eight of nine. Training set merge.	Pre-test for CB and BC relations for both sets. Post-testing after training each set of stimuli. Follow-up test. Test of set merge.	"while only one participant was able to meet criteria for demontration equivalence on the merge test, all participants consistently performed better as the test progressed, Which supports past research of delayed emergence of derived relations....During the sorting task however, all participants were able to meet criteria for demonstration equivalnece" (p.97)
14	Green	1990	AJMR	5 / Young adult females with mild mental retardation / 19-20	Two sets of stimuli. both: Two classes with three members. One set containing abstract forms. The other set with two classes where A sitmuli were spoken nonsense sylable and B and C stimuli were abstract forms.	Sorting was conducted with cards. Sorting: The 11th phase of the study to provide an additional indication to if the match-to-sample training and testing had produced classes of stimuli. Instruction: "put all the pictures that go together in stacks". (p.265)	Mimeographed pages (copied pages) to place the forms on marked areas, using pensil to mark sample and pensil to mark comparison when training in MTS format. OTM training. Sample stimuli displayed on topp of the page. Three comparisons displayed in a row on the bottom half of the page. Responding to sample with marking it with an X Praise as feedback when participants marked the correct comparison. "No, try another one" if participants marked an incorrect comparison. Blocks of 32 trials.	Mimeographed pages (copied pages) to place the forms on marked areas, using pencil to mark sample and pensil to mark comparison when testing in MTS format. Testing in phases: Equivalence - with 8 trials of each four untrained relations (BC and CB). Then symmetry relations - (BA and CA). No feedback.	"BC and CB relations emerged more quickly in the auditory-visual case. The results of the sorting test were also consistent with this findings" (p. 268).
15	Hove	2003	TPR	20 / College students / ?	One set of stimuli. Three classes and three members. Greek letters in black on white background.	Sorting was conducted with cards. Post-experimental task. The participants were asked to sort the nine cards.	Training conducted on a computer screen. Training in MTS format. Sampel stimulus on the left side of the screen. Comparison stimuli in six different positions on the right side of the screen. Two groups: One with OTM and one with MTO training Gradual introduction of three comparison stimuli, starting with only the correct, by touching the sample stimulus.	Testing conducted on a computer. Test block of 80 trials. (20 symmetry, 20 equivalence, 40 baseline trials). No reinforcement. Consequences for incorrect: 5-s. black screen and repetition of last trial. Stimuli presented both simultaneously and successively.	6 out of 10 participants in the OTM group sorted in accord with experimenter-defined classes. Four of these did not show equivalence during experiment. All 10 in the MTO group sorted in accord with experimenter-defined classes. Two of these did not show equivalence during experiment.
16	Lowe,Horne, Harris, and Randle	2002	JEAB	12 / Childern / 2-4	Two classes with three members Three abstract objects in the class of "vek" and three in the class of "zag" Two set of three everyday objects and 12 abstract objects of woden shapes was used.	The sorting described as *Category match to sample:* a test for emergence of categories after tact training. The children were asked to: "Look at this. Can you give Teddy the others?"	Children learned to tact two arbitrary sets of stimulus they were tested for arbitrary stimulus category afterwards. Those who failed with a Look-at-sample version - were tested again with a tact prior to selecting the other stimuli.	It was used a sorting test termed *category match to sample.*	" That the results of the three experiments are consistant with the naming account is as tru of the test failures as of the successes. The fact that severeal of the 12 subjects who learned the prerequisite arbitraty tact relations in Experiments 1a and 2 did not succeed on the category test until they were required to name the stimuli is what the account predicts." (p. 544) The authors say about the test that it "mark a substantial departure from standard techniques of paired stimulus selections". (p.546)

17	Mackay, Wilkinson, Farrell, and Serna	2011	JEAB	16 / Young adults / / ?	Three sets of stimuli one set for each condition (A,B,C). Under each condition: Four classes with three members, with the two A stimuli overlapping two classes. Abstracte stimuli as, trigrams as A stimuli, lexigrams as B stimuli and colored anlimal-like stimuli as C stimuli.	Conducted on a computer. Multiple-selection test performance. Used for documenting the precense or emergence of complex contextual control of stimulus class membership. A variant of MTS tasks where not only one single secection can be made. One, two or all three stimuli can be chosen under the presence of a sample.	Conducted on a computer. Training i MTS before experiment. A modified linear series training (first BC training and testing then AB training and testing) Reinforcement for correct choises: Chimes sounded. Under A and B conditions it was used different stimuli. Under C condition it was used same stimuli as in B condition and the training and testing order was changed.	Conducted on a computer. Interspersed in training. Testing each relation after training it.	"All participants demonstrated performances on symmetry and transitivity probes that were consistent with stimulus equivalence." (p. 97). "The multiple-selection test provides a highly efficient method for assessing the stimulus control of performances generated under conditions that establish stimulus classes that contain a member in common and thus could marge to form a single large class or intersect and remain as separate classes. " (p.98)
18	Nartey, Arntzen, Fields	2014	TPR	In experiment 1: 40 / university students / 18–36 In experiment 2 20 / university students / 20–35	Three classes with five members for each conditiion. Abstract symbols in black on white background were used in two conditions. Twelve abstract symbols and 3 pictures were used in two conditions. Hebrew and Arabic letters were used in discrimination training.	Sorting was conducted with the same stimuli as printed copies. Sets of plastic laminated cards were used, one set for each condition. Sorting test before training of class formation and after MTS based class formation test. Asked to: "put them into groups" (p.779)	Conducted on a computer. Two experiment both with 4 conditions (only abstract stimuli A-as-SD, only abstract stimuli E-as-SD, abstract stimuli pluss three pictures A-as-picture, abstract stimuli pluss three pictures E-as-pictures) MTS based training with linear series structure. In experiment 1: baseline relations established in a simultaneous protocol and serialized training. In experiment 2: Baseline relations established in concurrently training. Mastery criterion 90% before next block trained. Maintenance of baseline relations: In mixed trials with fading of programmed feedback like 75%, 550%, 25% and 0%.	Conducted on a computer. Two test blocks of 90 trials a total of (36 baseline trials, 36 symmetry trials, 54 one-node trials and 36 two.node trials and 18 three-node trials). Presented randomly. No feedback Mastery criteria defined as 90% of trials for each type of relation.	"For 92% of the participants, then, the yields computed from the sorting data corrosponded exately to the outcomes produced from the derived relations test." (p. 784)
19	* Nartey, Arntzen, Fields	2015	LB	50/ University students / 19-23	Three sets of stimuli: 15 stimuli in each set organised in three classes with five members. Pluss nine extra stimuli used for discrimination training of the abstract C stimuli. Abstract symbols in black on white background used for two groups and C stimuli as meaningful stimuli in one group.	Sorting conducted with the same stimuli as printed copies All groups, but the group with C stimuli as meningful pictures started the experiment with a pre-class formation sorting task, with laminated cards. After participants attempt to equivalence class formation, the participants did a post-class formation sorting test. Iinstructions were: "put the cards into groups"(p.22)	Training conducted on a computer Five groups. (1. only abstract stimuli, 2. meaningful C stimulus, 3. abstract C stimulus pretrained with simultaneous and succesive discrimination, 4. Abstract Cstimulus pretrained with only simultaneous discrimination, 5. Anstract C stimulus pretrained with only successive discrimination.) Linear series training struccture in MTS format. Four serializes trials phases, with trial types presentes in a random order afterwards a mixed trials phase with trials presented randomly. The sample in the middle of the screen and three comparisons in the corners of the scren. A response to sample. Programmed consequences. Blocks of 36 trials. Blocks repeated until correct response on at least 90% of trials of each type. Maintenance of baseline relations: The percentage of trials with feedbask was redused to 75%, 50%, 25% and 0 % before testing	Testing conducted on a computer Two test blocks of 90 trials a total of (36 baseline trials, 36 symmetry trials, 54 one-node trials and 36 two.node trials and 18 three-node trials). Presented randomly. No feedback. Mastery criteria defined as 90% of trials for each type of relation.	"This shows that performance on card sorting was influenced by the contingencies set in the training and testing for derived relations. These data also show that class-based behavior generalized between two trial formats; MTS trilas during class formation, and sorting during post class-formation testing" (p.30) "Participants who did not form classes also did not sort stimuli into experimenter-defined classes." (p.27)

20	Nedelcu, Fields & Arntzen	2015	JEAB	48/undergraduate /18–50	Two classes of five members Four conditions. All classes consisting of abstract nonsense syllables of consonant-vowel-consonant combinations. Exept one condition in which the classes included four nonsens syllables and one familiar pictorial stimulus	Sorting conducted with cards. Only conducted with six participants from one group. The cards were shuffeled so the ordering of the cards in the deck was not systematic. Instruction: "please sort the cards into stacks of related words" (p.355) "To assess the generalization of class-consistent responding from the matching-to-sample format to a different stimulus presentation format".(p.354)	Training conducted on a computer. Four groups with different variables of conditional discrimination in preliminary training. Equivalence-class formation training of baseline relations in MTS format. Simultaneous protocol. Baseline relations established serially in individual blocks. One sample, three comparison stimuli in each trial. Each block repeated until 100% accuracy. Maintenance of baseline relations: Fading the programmed feedback like 75%, 25%, 0% . mastery criterion 94 % of trials - if failed to meet criterion the block would be repeated up to five times.	Conducted on a computer. Series of six 20-trial blocks. (four baseline, 16 derived relations) presented in randomized sequences. No feedback. Mastery criterion to define class formation was 90%.	Three of six participants who did the sorting test showed maitenance of experimenter-defined classes and showed merging of stimuli trained to C stimulus. One of the six participants did not show class formation in the test for emergent relations - and did not sort in accord with experimenter-defined classes. Two participants did not show full merge of the stimuli into one class in the sorting test.
21	Pilgrim and Galizio	1996	Cb	47 / College students / ? /	Two classes with four members 8 Abstract 3-dimensional objects	One experimental group and the control group sorted all stimuli, Another experimental group sorted onlly 6 of the stimuli (B, C, D). Instructions: "Please place these objects into groups whatever groups you think are most appropriate" (p.188)	Two experimental groups + one control group. The experimental groups given the same training control group no training. 3 two-choice conditional discrimination (AB, BC, and AD) Mastery criterion 14 of 16 trials correct for two consecutive blocks.	Sorting test. No other probes given.	"Nine of the 16 subjects in each experimental conditon sorted stimuli into groupings consistent with the equivalence classes that would be predicted" "none of the 15 control subjects sorted stimuli into class-consistent groupings" Significant differences between experimental and control conditions. "Sorting might prove a useful alternative to consider when questions arise concerning an active or instructive role for probe trials in the emergence of equivalence relations.(p. 188)
22	Sigurdardottir, Mackay, and Green	2012	JEAB	4 / Typical adults / 26-54	Six auditory-visual classes with three members. Each consisting of a singular Icelandic spoken noun, a corresponding printed word and a corresponding picture. In second experiment: Stimuli is extended with six auditory-visual classes with three members. Each containing the same words as in the first experiment but now in plural. Pluss two contextual stimuli, Icelandic words for "number" and "gender".	Sorting conducted as a tabletop sorting task Sorting used in Second experiment to demonstrate contextual control. Instructions to participants: "Aarrange the cards in your hand in any way you want depending on wich cards you think go together when you have this card (pointing at the contextual sitmulus) in front of you" (p. 16)	Training conducted on a computer. In first experiment: First identity matching test. Then, spoken word-printed word (AB) conditional relation traind. At last the word-picture conditional relation were trained (AC). MTS based training at min. 216 trials. A response to sample needed, for comparisons to appear. AB and AC training each had 36 trials. Programmed consequences as a jingle and flashing light for correct responses. Training for maintenance. Block with 50% programmed consequences and if correct responses maintained 0% programmed consequences. In second experiment: Training for contextually controlled conditional relations. In the presens of the Icelandic word for "number" and "gender." In a two-choice MTS.	Testing conducted on a computer In first experimet: Tested the BC and CB relatins. No programmed consequences. Mastery criterion was 35/36 trials. If criterion was not met, training and testing was repeated.	"Contextually controlled classification was also demonstrated by all participants on a sorting test, though some participants required reviews of the contextual control training and retesting before they produced that outcome" (p. 22)

23	Smeets, Dermot, Barnes-Holmes	2005	TPR	16/ Children/ 5	Three classes of 5 members Four sets of stimuli: Tree sets consisted of color patches, geometrical forms, letters, scribble-typeconfigurations and schematic drawings of familiar abjects. These stimuli were used for pre-training. The fourth set consisted of abstract forms used during the experimental phases pluss auditory stimuli dictated by the experimenter.	Pretraining sorting as phase 1 Sorting test as phase 11 and 19 Sorting were conducted on the table with cards. The experimenter placed an unorganized pile of nine card on the table. Instruction: "Make piles of pictures that go together"(p. 493)	Training identity matching as phase 2 Baseline training as phase 4–6 and 13–15 Training on table, stimuli presented on laminated white cards. Comparisons near the bottom of the card and sample on the top of the card. Training in MTS format. Correct responses followed by praise and the delivery of a bead. Incorrect responses were followed by "Wrong! No bead" (P.488). Mastery criterion: N-1 trials correct in one block or N-2 trials in each of two consecutive blocks of 18 trials.	Testing on table Testing reflexivity as Phase 3 and 12. Baseline test as phase 7 and 16 Equivalence test as phase 8 and 17. Symmetry test as phase 9 Naming test as phasae 10 and 18 No feedback after trials, but after completion of a block 12 beads were given.	" Class consistent sorting was seen 18 times, once after failing the equivalence test .....and 17 times after passing that test" (p.497)
24	Smeets, Dymond, and Barnes-Holmes	2000	TPR	1. Experiment: 23 / Third-year psychology students/ ? 2. Experiment: 120 / First year psychology students /? 3. Experiment: 77 / First year psychology students / ?	In all experiments under all conditions. Two classes of 3 members. Six symbols served as experimental stimuli.	Sorting as pensil and paper procedure. In experiment two, the experiment also examined the symmetry and symmetric transitivity with sorting tests. This was done on a page, with "the instruction to categorize the sitmuli into two groups." (p.347) In experiment three: As in experiment two but with the instruction: "you may skip problems that you find impossible to solve".(p.348)	Training as pencil and paper procedure. Three experiments. First experiment had two conditions. First condition: Linear series training. (A-B, B-C) Second condition: (A-B, C-D) training. First the participants were trained in MTS by identity MTS. Afterwards they got rules about relations between the stimuli to memorize and reheardh, with 40 training trails (10 of each relation) and with the ruels present, Second experiment had three conditions: The conditions were different in the order the stimulus relations were tested. Third Experimient: As experiment two but with instruction.	Testing as pencil and paper procedure. First experiment: After training participants got 32 test trials (16 baseline and 16 C-A relations). Second experiment: As first experiment but also sorting tests. Third experiments : As second experiment but with instruction: "you may skip problems that you find impossible to solve"	"These findings may indicate that, at least for the test that was used here, class-consistent sorting does not provide convergent validity for stimulus equivalence." .... "Subjects who responded correctly on only symmetry or symmetric transitivity tests did not meet the criteria for stimulus equivalence. Yet many of these subjects (61%) sorted the stimuli in a class-contingent fashion." (p. 352)
25	* Varelas and Fields	2017	TPR	55 / College students / ? 32in experiment, 23 as controle group	Three classes of five members. All stimuli describing three stages of prenatal development. A: Stimuli the names B: Stimuli the time C: Stimuli defining characteristic	Sorting conducted with cards A randomized shuffled deck of cards given to participants Instruction to " ...sort them into groups on your desk. Organize the cards however you belive the information on them goes together.You may create howeveras many groups as you oike. there is no penalty for any wrong answer." (p.73) Initial sorting test and post-training sorting test.	Training conducted with a clicker device. Feedback were given to all students at the same time showing the percentage of students who selected each of the answers. The correct answer was also shown. MTO training structure. AB relations trained to 90% of students answered correctly. Then CB trained to 90% of students answred correctly. Maintenance of relations Mixed training until 90% of students answered correctly on 50% feedback then mixed training unti 90% answered correctly with no feedback.	A post- training sorting test were used as a test for immediate emergence of equivalence classes:	"Sorting documented the immediate emergence of equivalence classes by 26 of 32 participants."(p.78)

Note. In the matrix is used the following abbreviations: TPR=The Psychological Record, JABA=Journal of Applied Behavior Analysis, EJOBA=Eueopean Journal of Behavior Analysis, JEAB= Journal of the Experimental Analysis of Behavior, AJMR=American Journal on Mental Retardation, LB=Learning and Behavior, Cb= Chapter in book.

Table 2

Journals Publishing Studies

Journal	Number of studies	Percentage of sample
The Psychological Record	11	44%
European Journal of Behavior Analysis	5	20%
Journal of the Experimental Analysis of Behavior	5	20%
Journal of Applied Behavior Analysis	1	4%
American Journal on Mental Retardation	1	4%
Learning and Behavior	1	4%
(Chapter in book	1	4%)

Table 3

Population Characteristics

Characteristics	Number of studies	Percentage of sample
Age:		
Children (Age: 2–5)	2	8%
Adult age (Age: 18-62)	23	92%
Other population details:		
With disability	2	8%
Undergraduate, college or university students	15	60%
Not specified	8	32%

Table 4

Stimuli Used in the Studies

Type of stimuli	Number of studies	Percentage of sample
Abstract stimuli black on white background	7	28%
Abstract stimuli black on white background + meaningful stimuli	4	16%
Greek and Arabic letters pictures and nonsense syllables	1	4%
Dictated names, photos of faces, Written names, photos of nameplates	1	4%
Pictures of students	1	4%
Phonologically correct non-words as printed and spoken	1	4%
Pictures of patterns + olfactory stimuli.	1	4%
Abstract stimuli black on white background + spoken nonsense syllables	1	4%
Greek letters in black on white background	1	4%
Abstract objects, every days objects, nonsense syllables	1	4%
Abstract nonsense syllables + familiar picture	1	4%
Abstract objects	1	4%
Icelandic spoken noun, printed word And picture	1	4%
Symbols (math and special characters)	1	4%
Three stages of prenatal development	1	4%
Abstract forms+ auditory stimuli	1	4%

Note. Stimuli used in pre-experimental training is not included

Table 5

Number of Classes and Members in Classes

Number of classes	Number of members	Number of studies	Percentage of sample
2	3	3	12%
2	4	1	4%
2	5	1	4%
2	7	1	4%
3	3	4	16%
3	4	1*	4%
3	5	11	44%
4	3	2	8%
6	3	1	4%

Note: The symbol * marks that in one article (Dickins, 2015) it was used different number of classes and members in three experiments. We have only included numbers of classes and members from the third experiment, because this is the experiment and the results referred to in this article.

Table 6

Variables in Procedures for Performing Sorting Tests

Sorting conducted:	Number of studies	Percentage of sample
On top of a table with cards	17	68%
On a computer screen with cards	5	20%
Giving objects to a teddy bear	1	4%
Grouping objects on a table	1	4%
Using pencil and paper	1	4%

Table 7

Instructions Used Before the Sorting Test

Instructions	Number of studies	Percentage of sample
Put them (stimuli) into groups (4, 7, 11, 12, 18. 19, 21)	7	28%
Categorize the stimuli (1, 2, 9, 10)	4	16%
Sort the cards into stacks (14, 20)	2	8%
Put all the cards/pictures that go together in stacks (6)	1	4%
Place the objects in three piles (13)	1	4%
Put these into groups as you feel like. (And "participants were informed that the top card had to be moved to a different location on the screen that was close to other related stimuli so that they formed a cluster that was separated from other clusters that contained stimuli from different sets.") (5)	1	4%
Put them together in the way that you think is correct, but make sure that all stimuli are visible. (3)	1	4%
Put them into some kind of order (8)	1	4%
Sort the cards (15)	1	4%
Look at this. Can you give teddy the others?(As a *category match to sample*) (16)	1	4%
Make piles of pictures that go together (23)	1	4%
Categorize into groups (24)	1	4%
Sort them into groups on your desk. Organize the cards however you believe the information on them goes together. (25)	1	4%
Not specified (17, 22)	2	8%

Note. Numbers in parenthesis refers to numbers on reviewed articles in Table 1

Attending Behavior in the Presence of Compound Stimuli

Abstract

In the present experiment, we employed Chinese characters as compound stimuli to investigate the variables influencing attending behavior in matching-to-sample (MTS) and sorting tests. The Chinese characters were used with the radical and without the radical. Twenty participants were randomly assigned to two different sequences of experimental conditions, and the participants in both sequences were exposed to pre-sorting of stimuli, first without and then with the radical, and subsequently exposed to a hybrid MTS (H-MTS) training in which the compound stimuli included identical radicals and therefore were only partly arbitrary. After training the baseline relations, the participants that were assigned to Sequence 1 had the sorting tests of stimuli, first without and then with the radical, followed by a H-MTS test, including stimuli with the radical, and finally the post-sorting of the stimuli without and with the radical. The participants assigned to Sequence 2 had a H-MTS test including stimuli with the radical, then sorting tests of stimuli without and with the radical, followed by a H-MTS test including stimuli with the radical. The immediate emergence of all stimulus classes of the compound stimuli with radicals was seen in seven of the ten participants in both sequences, whereas for the compound stimuli without radicals, it was only seen in two of the ten participants in Sequence 1 and one of the ten in Sequence 2.

Keywords: attending behavior, class formation, sorting, matching-to-sample

Humans are often found in situations in which many stimuli impinge simultaneously on their behavior, and many of these stimuli are compound stimuli. Examples of situations in which compound stimuli affect behavior are when one is switching back and forth between listening to the drums and the piano in a piece of music and when one is looking for the car keys while looking at them. We would colloquially say that we are "paying attention" or "not paying attention". Investigations of this type of behavior in cognitive psychology are often done by making models, for instance, of different types of attention capacity allocation (Mcleod, 1977) or of attention systems as a complex network of interconnected subsystems (Garon, Bryson, & Smith, 2008). In behavior analysis, there is another approach. Skinner (1953) explains that the criterion of attending is not about how we use our eyes or ears but rather if the stimulus has any effect upon our behavior. He argued that attending is not a form of behavior but rather "a controlling relation—the relation between a response and a discriminative stimulus"(Skinner, 1953, p.123). Nontechnical terms such as attention, paying attention and not paying attention are used to describe differences in the behavior of an organism responding to some stimulus properties and not to others. Rather than these nontechnical terms, it is better to use the verb "attending" or to speak about attending behavior to emphasize that the differences in behavior are due to differences in processes and not an unavailable "thing" called "attention" (Donahoe & Palmer, 1994). Catania (2013) says that "attending is not defined by movement; it is defined in terms of its consequences" and it is "appropriate to talk about attending as an operant" (Catania, 2013, p. 141).

Reynolds (1961) investigated attending behavior in an experiment with two pigeons trained to key peck on two compound stimuli, a triangle-on-red stimulus correlated with reinforcement and a circle-on-green stimulus correlated with extinction. In a subsequent attending test (i.e., test under extinction, without consequences and with the two compound stimuli split up into four stimuli) one pigeon mainly attended to the form of the previous reinforced stimulus (i.e., had most pecks on the triangle), while the other attended mainly to the color of the previous reinforced stimulus (i.e., had the most pecks on the red color). Even though it showed that the pigeons under extinction pecked on either the form or the color of the previous reinforced compound stimulus, it also showed that the birds did not generalize between the stimuli. The pigeons discriminated almost solely in one dimension of the compound stimuli and not in the other; we could say they were attending only in one dimension and failed to attend in the other. Instead of talking about the dimensions of a compound stimulus as being salient, (e.g., in Reynolds' (1961) experiment, the triangle or the red color as being salient properties), it is better described with Catania (2013) words: "salience isn't a property of a stimulus; it is actually a property of the organism's behavior with respect to that stimulus" (p. 140).

Donahoe and Palmer (1994) described four reasons why attending behavior can fail to occur in an environment: First, if a present discriminative stimulus is not observed or sensed by the organism. Second, if the history of selection by consequences with respect to the stimulus failed to bring the behavior under the control of the stimulus in that environment; for instance, when a learner fails to attend to a second stimulus introduced after the first stimuli has gained stimulus control, it is called blocking. Third, if a stimulus has gained stimulus control in a context but fails to function as a discriminative stimulus outside this context. Fourth, if the simultaneous occurrence of discriminative stimuli interferes with the responses that are normally emitted in the presence of those stimuli. Donahoe and Palmer (1994) list different outcomes of the last situation: that responses occur successively, that a type of mixture of the responses occurs or that only one of the responses occurs.

In the literature, we find several definitions of compound stimuli used in different experiments. For example in experiments on respondent conditioning, conditioned stimuli are presented together and obtain the capacity to evoke a conditioned response.(Augustson, Dougher, & Markham, 2000; Recorla & Wagner, 1972) In experiments on operant conditioning, compound stimuli are often used in discrimination training. Reynolds (1961) used two separate stimulus elements exposed upon each other in the discrimination training and later separated those elements when testing the discriminative stimulus control. Debert, Matos, and McIlvane (2007) suggest a definition of compound stimuli to "include stimulus elements joined temporally or spatially – components that could be separated and recombined without loss of discriminative control" (p. 90). In the present study, we define a compound stimulus to be a complex stimulus consisting of two or more individual stimuli presented simultaneously, and we, too, will separate the parts of the compound stimulus to perform investigations.

Attending to compound stimuli in humans was addressed by Clark L. Hull (1920). He investigated the idea that concepts are the descriptions of classes that are defined by the presence of all of a set of attributes. With this approach, he launched the idea of concept learning as discrimination between relevant and irrelevant attributes. He studied this by letting participants learn a separate nonsense name for each member of sets of twelve Chinese ideographs. The name was related to a specific radical (Chinese characters are built of components called radicals) in the Chinese character. The participant had to discriminate the characters and learn all of the twelve nonsense names in one set before they were exposed to a new different set but with identical radicals and the same names correlated with the radicals. The participants were expected to tell the name of each character before being prompted. As consequences of their behavior, the participants were told if the responses were correct or incorrect, and, as Hull said, "If he could not react correctly, he had the annoyance of failure, and if he succeeded, he had the satisfaction of conscious success" (p. 14). We would say that discrimination was learned as a result of its consequences. Hull found that each participant's performance of correct responses as a percentage went up progressively for each set of new characters. After five such sets, the participants would name more than half of the characters correctly when seeing them for the first time and without being able to tell what the common radical was or to sketch it. In the experiment of Hull (1920), he illustrated how stimulus classes can be generated. However, it is not possible to infer whether stimulus equivalence, as defined by Sidman and Tailby (1982), had emerged.

In the field of stimulus equivalence, studies in concept learning and other complex behavior such as language and remembering have, in the last decades, often been studied by the use of conditional discrimination procedures in a matching-to-sample (MTS) format. In this procedure, using arbitrary stimuli, the participants are put in the presence of a sample stimulus and a fixed number of comparison stimuli and taught to choose the experimenter-defined stimulus (e.g., to choose B1 in the presence of A1, not B2 or B3 and to choose C1 in the presence of B1, not C2 or C3). The MTS format is often used for both the training of baseline relations and subsequently to test for untrained emergent relations and document class formation. The untrained responding on the test for emergent relations must have the properties of reflexivity, symmetry, and transitivity to qualify as stimulus equivalence. An equivalence class contains a fixed number of disparate stimuli that are related to each other in an interchangeable manner, and stimulus equivalence is defined as stimulus substitution (Green & Saunders, 1998).

Arntzen (2012) describes how the training structure in MTS can be varied: In a *linear series* (LS) training structure for three 3-member classes, first all AB relations are trained and then all of the BC relations. In a *many-to-one* (MTO) training structure, all the AC and BC relations are trained, and in a *one-to-many* (OTM) training structure, it is the AB and AC relations being trained. Originally, Sidman and Tailby (1982) suggested that the outcomes in the test for emergent relations should not vary between the different training structures, neither in order or direction. At present, there is no unambiguous evidence of which of the training structures, MTO or OTM, is the most effective, but the LS training structure has been shown to be the least effective considering the production of stimulus equivalence in a simultaneous protocol (Arntzen, Grondahl, & Eilifsen, 2010). Arntzen et al. (2010) argued on the basis of studies that “the MTO training structure is a very effective structure in the sense that it will give higher yields in the structures that follow ” (Arntzen et al., 2010, p. 457). In the present study, the MTO training structure was selected before OTM or LS to give high yields in the structures that follow after the MTS training.

Dymond and Rehfeldt (2001) have proposed the use of measures other than the percentage of correct responding when investigating derived stimulus relations and the emergence of relations that are not directly trained. They argued, among others, for the use of the reaction time to the comparison stimuli and the sorting of stimuli as additional measures. The reaction time to the comparison stimuli in testing trials has been examined in some studies (e.g., Arntzen, Braaten, Lian, & Eilifsen, 2011; Arntzen, Galaen, & Halvorsen, 2007; Arntzen et al., 2010; Arntzen & Hansen, 2011; Bentall, Jones, & Dickins, 1998; Eilifsen & Arntzen, 2009; Spencer & Chase, 1996). There is shown to be an increase from the mean of the median reaction times in the last 5 trials in training to the first five trials in testing the baseline relations. There was also found to be an increase in the mean of the median reaction times from the trials testing the baseline relations to the trials testing the symmetry relations and again from the trials testing the symmetry relations to the trials testing the transitivity and global equivalence relations. In these studies, there was also found to be, on average, a decrease in the mean median reaction time in the last five trials compared to the mean of the medians in the first five trials in all relations. We will investigate the reaction time in this study and attempt to replicate these findings.

As mentioned above, (Dymond & Rehfeldt, 2001) also suggested the use of sorting tests to document equivalence class formation. Recently, experiments have implied that stimulus sorting tests can be used as a measure of the presence and/or emergence of equivalence classes (e.g., Arntzen, Granmo, & Fields, 2017; Arntzen, Norbom, & Fields, 2015). In Arntzen et al. (2015), the sorting test was conducted immediately after the training of the baseline relations (TBR) and was then followed by an MTS equivalence class formation test to assess if the emergence of all class-based relations had occurred. In this experiment, 100% of the participants, who previously had demonstrated the formation of equivalence classes, showed the immediate emergence of stimulus classes by sorting. In Arntzen et al. (2017), they asked whether this finding was a general phenomenon and whether the yields would be the same with participants that had no prior experience in forming equivalence classes? This question was addressed by the design of an experiment on two groups with different placements of the MTS test in the phases of TBR and sorting. In Group 1, with sorting following the TBR, 50% of the participants showed the immediate emergence of stimulus classes, and in Group 2, with sorting following the MTS-test, 30% of the participants showed the emergence of the three classes. They concluded that the percentage of participants showing the immediate emergence of stimulus classes by sorting was found to be lower than that in the previous study.

In the present study, we used Chinese characters with common radicals as compound stimuli like Hull (1920) did. We asked if MTS training with a set of stimuli having some identical features (i.e., the radicals) would have any impact on the results when sorting the same set of stimuli but without the identical features, making them arbitrary stimuli. We used a hybrid-MTS (H-MTS) procedure that included a mix of identity and arbitrary aspects of stimuli as Chinese characters having identical radicals. The experiment was arranged to study sorting and the immediate emergence of stimulus classes after training baseline relations in accordance with the findings of Arntzen et al. (2015) and Arntzen et al. (2017). The experiment attempted to replicate the findings on the changes in reaction time from baseline training to testing.

Method

Participants

Twenty participants (17 women and 3 men, $M_{age} =$ 26,95 years, age range: 20-50 years), with education levels from undergraduate to graduate, were recruited by personal contact. No compensation was offered for participation. Before the experiment started, all the participants read an information sheet, but the purpose of the experiment was not mentioned. Translated into English, the text stated as follows:

> The project is in the field of learning psychology and aims to provide more insight into the variables that affect how categorization emerges. Providing knowledge about variables that influence the formation of stimulus classes can be essential to increase the understanding of the phenomena people usually call memory, problem solving, language and the use of symbols. To not affect the research results, I cannot go into detail or explain further what this means before the trial starts. However, all participants can see their results and obtain an explanation of categorization and stimulus equivalence as research fields and what the present research specifically examines after they have participated. During the debriefing, there will also be opportunities to ask questions.

Their rights as research participants were explained. The participants were assured anonymity and were informed of their right to withdraw from participation at any time without any negative consequences. The participants were also informed that no harmful effects were expected in the present experiment. Each participant signed an informed consent form. The participants were informed that the experiment would last from one to two hours. After having finished the experimental session, all participants were debriefed and presented with their results from the sorting and MTS tests. The participants did not have any knowledge of or experience with this type of experiment or with stimulus equivalence as a discipline.

Apparatus and Setting

Apparatus. All tasks in the present experiment were presented on a computer screen. The experiment was run on a Hewlett-Packard HP EliteBook 876w PC with the Windows 7 Professional 32-bit operating system. The processor was an Intel® Core™ i5-2540M CPU @ 2,60 GHz. The display of the laptop was a 1.LG T1710 on a standard VGA Graphics Adapter with a resolution of 1280 x 1024, and the screen was 38 cm wide by 21 cm high. A wired mouse with 3500 dpi precision and a 3,5G infrared sensor was used. Custom-made software was used to run the MTS training and testing, and tailor-made software was used for the sorting test.

Setting. The setting was two similar rooms (approximately 3 x 3 meters), both in quiet surroundings. Unnecessary items were removed or covered with white tablecloths, and items that could be used for help (like pens and paper) were removed. The participants were asked to leave their mobile phones outside the room. The participants were placed sitting at a desk (0,50 x 1 meters) in the corner of the room, facing the computer and a monochrome wall. On the participant's side was a covered window, and the experimenter sat outside the door to the experiment room.

Design

We applied a within-subject design in combination with a group study to explore the research questions. The experiment was done with nine direct replications under each condition (i.e., sequence). The two different sequences were arranged to control for order effects (see Table 1). The participants were assigned randomly to the two sequences, and statistical analyses were completed between the sequences. All participants began the experiment with a pre-class formation-sorting test without radicals (Pre-SRT-WoR), followed by a pre-class formation-sorting test with radicals (Pre-SRT-WR). This was to determine if the experimenter-defined classes were formed before we started the training of the baseline relations. If all of the three experimenter-defined classes were formed in the Pre-SRT-WoR, this would be a reason for exclusion, but not for Pre-SRT-WR. The participants completed the training of the baseline relations with radicals (TBR-WR) in the H-MTS training phase. For the participants assigned to Sequence 1, the TBR-WR was followed by sorting the stimuli, first without (SRT-WoR) and then with (SRT-WR) the radical and afterwards the H-MTS test with radicals (H-MTS- WR-test). Finally, the experimental session was ended by the sorting tests, first without (Post-SRT-WoR) and then with the radical (Post-SRT-WR). The participants assigned to Sequence 2 were exposed to an H-MTS test with radicals (H-MTS-WR-test-1) immediately after the TBR-WR, followed by post-sorting tests without and with radicals, and the experiment was finally ended by a second H-MTS test (H-MTS-WR-test-2).

Dependent and independent variables.

The dependent variables in the present experiment were the behavior of using the mouse to click on the comparison stimulus in the presence of the sample stimulus. Different measures were used to examine the changes in the dependent variable, for instance, numbers of correct and incorrect trials, reaction time and class formation. The independent variable was the programmed consequences meant to differentially reinforce the choice of what the experimenter had defined as the right comparison stimulus

Stimuli

Two sets of stimuli, all Chinese characters, were used. Each set consisted of nine stimuli, with three stimuli in each class. As we see in Figure 1, one set contained stimuli with radicals (WR) (i.e., identical features) in each class, while the other set was the exact same set of Chinese characters but without the radicals (WoR).

Procedure

Sorting. The stimuli in the sorting test were the two sets of three 3-member classes shown in Figure 1. During the sorting tasks, no feedback or programmed consequences were provided. The sorting test started with only one of the stimuli displayed on the computer screen, while the rest of the stimuli were hidden behind this stimulus. Before the sorting tests, the participants were asked to read the following text (in Norwegian) on the computer screen:

> "You will now see a pile of pictures, and you are going to drag all the pictures out by using the left button on the mouse. Press the button on top of the upper picture and drag it to the side, and do the same for all nine pictures. Now, you are going to sort the pictures and mark how you have sorted them. The marking is done by holding down the left button on the mouse and dragging the cursor over the screen."

The first four participants only received the two last sentences of this information on the computer screen and the rest of the information verbally. Before leaving the room, the experimenter told the participant that the sorting task should be completed two times and that the participant had to press the button labeled "finished" when done with the task. The experimenter also told the participant to call the experimenter when the task was finished. When the participant pressed the button "finished", the computer program would ask, "Are the pictures sorted and a marking of the sorting done?", and the participant had to click on the button "yes", or otherwise the task would continue. When "yes" was answered, a screenshot was automatically recorded and saved on the computer.

Conditional discrimination training. After the pre-class formation-sorting phase, all participants underwent conditional discrimination training (i.e., with the partly identical stimuli). The following text in Norwegian was presented on the screen:

> "A stimulus will appear in the middle of the screen. Click on this using the computer mouse. Then, three other stimuli will appear. Choose one of these using the

computer mouse. If you choose the stimulus we have defined as correct, words such as 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. 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 so far, you can get all of the tasks correct. Please do your best to get everything right. Good luck!"

In addition, the experimenter told the participant to call the experimenter when the task was finished.

The conditional discrimination training was done with the baseline relations of stimuli WR and was thus an H-MTS task. The baseline relations (AC, BC) were established in a simultaneous MTS format, and baseline trials were presented concurrently using an MTO structure. The intertrial interval was 1000 ms, and the interval with programmed consequences was 500 ms. In each training trial, the sample stimulus was presented in the middle of the screen, and a mouse click on the stimulus was followed by three comparison stimuli that were presented in random corners from trial to trial. A click on the correct comparison stimulus resulted in a programmed consequence such as the text stimulus "fantastic" or "correct" presented on the screen, while an incorrect choice resulted in the programmed consequence "wrong" as a text stimulus on the screen. Each trial type (e.g., A1/C1, B1/C1) was presented five times in a block. Hence, each block consisted of 30 trials. When 100% of the baseline relations in a block were acquired with the programmed consequences for every correct choice, the probability of the programmed consequences was changed from 100% to 50% and finally 0% of the trials in the block. The training of the baseline relations ended when a participant responded accurately on 100% of the trials in a block with a 0% probability of programmed consequences.

Testing for emergent relations. In the test block, experimenter-defined responding on the baseline relations and the emergent relations of equivalence classes were assessed. The baseline relations (AC, BC), symmetry (CA, CB) and transitivity/equivalence (AB, BA) were tested. The testing was done in a mixed test block with no programmed consequences. Each trial type was presented three times, making 18 baseline trials, 18 symmetry trials and 18 transitivity/equivalence trials, for a total of 54 trials. The criterion for responding in accordance with the equivalence was 95% correct responses for all trial types.

Criteria for interrupting the experiment. If the participant asked the experimenter to stop, the experimenter would immediately stop the experiment.

Dropouts and Remarks

After approximately 32 minutes in the training of the baseline relations, one participant told the experimenter to stop the experiment. One participant informed the experimenter during the debriefing that she was studying Chinese and was for that reason withdrawn from the results. Two participants were withdrawn from the results because of programming errors. Participant numbers 15621, 15622, 15623, and 15624 served as replacements for these participants.

Reliability

To make sure that the results of the sorting tests were reliable, two observers assessed the screenshots of the sorting tests. Their Inter Observer Agreement (IOA) was measured as percentage of agreement (i.e., the number of sorting trials with total agreement on the distribution of stimuli divided by the total number of sorting trials then multiplied with 100).

Statistical Analysis

Statistical analysis was conducted. Independent-sample t-tests were used to compare the two sequences with respect to the number of baseline training trials used. A chi-square test was used to compare the results on mastery of all the three experimenter-defined classes or not, from the pre-class formation sorting tests to the post-class formation sorting tests and for stimuli WR and stimuli WoR. Independent sample t-tests were again used to analyze the differences in the medians from the first five to the last five trials in all trial types within all three H-MTS tests. Finally, an analysis of variance (ANOVA) was used on the median reaction times from the first five and last five trials of each trial type to compare the reaction times between the H-MTS tests.

Results

Analysis of Sorting Data

Table 2 presents the performance of each participant. The performance was scored as Y(i.e., yes) or N (i.e., no) depending on if a participant formed the experimenter-defined equivalence classes or not in the H-MTS-test. The three-digit strings under the headlines indicate the representative stimuli in the marked groups (i.e., groups of stimuli with a line drawn around them) produced in the sorting test. For example, as shown in Figure 2, participant 15623 in the Pre-SRT-WoR sorted the stimuli into three marked groups. The groups of sorting is transcribed as 210 120 003 in Table 2. The first cluster (210) means that there were two stimuli from Class 1, one stimulus from Class 2 and no stimuli from Class 3. In the second cluster (120), there were one stimulus from Class 1, two stimuli from Class 2 and no stimuli from Class 3, and in the third cluster (003), there were no stimuli from Class 1, no stimuli from Class 2 and three stimuli from Class 3. The experimenter-defined classes produced by the participant are illustrated in **bold** font. A full string with experimenter-defined classes will look like **300 030 003.**

Pre-Class Formation Sorting Tests in Both Sequences

Pre-SRT-WoR tests in both sequences. No participants sorted the stimuli in accordance with all the experimenter-defined classes. In Table 2, the strings under Pre-SRT-WoR show that the participants marked from one to five groups when sorting the stimuli. The number of stimuli in the marked groups varied from one to nine, and they contained stimuli in a mix from one, two, or all three of the experimenter-defined classes. Three participants (15623, 15622, and 15603) sorted Class 3 corresponding to the experimenter-defined classes. Participant 15610 marked all the stimuli in one group, as shown by the cluster 333 in Table 2. Between the sequences we see two participants form Sequence 1 and one from Sequence 2 sorted the Class 3 corresponding to the experimenter-defined class. We see that the participants in Sequence 1 marked from one to four groups of stimuli. Wheras, in Sequence 2 the participants marked from two to five groups.

Pre-SRT-WR tests in both sequences. Two participants (15612 and 15619) out of 20 sorted all three experimenter-defined 3-member classes, and the rest of the participants sorted the stimuli into one to four marked groups. As shown in Table 2, one participant (15621) sorted the stimuli in Classes 1 and 2 in accordance with the experimenter-defined 3-member classes, and three participants (15622, 15618 and 15624) sorted one of the experimenter-defined classes. Participant 15610 marked all 15 stimuli together, and this is transcribed as 333 in Table 2 and shown in the screenshot in Figure 2**.** Between the sequences we see in Table 2 that no participants in Sequence 1 sorted all the three 3-member experimenter-defined classes. But, two participants sorted all the experimenter-defined classes in Sequence 2. In Sequence 1 two participants sorted two single classes corresponding to the defined classes. In Sequence 2 nine single classes were sorted corresponding the experimenter-defined classes. In Sequence 1 the participants marked from one to four groups of stimuli and in Sequence 2 the participants marked from two to four groups of stimuli.

Consistent responding in pre-class formation sorting tests. Participant 15610 did the exact same sorting of stimuli in Pre-SRT-WR as in Pre-SRT-WoR, marking all the stimuli into one group in both sorting tests (see Figure 2). Except for one stimulus, Participant 15615 marked the stimuli in the same groups in Pre-SRT-WR as in Pre-SRT-WoR. We see in Table 4 that the sorting from Pre-SRT-WoR to Pre-SRT-WR only differs by stimulus A1 being placed in a different group. None of the participants sorting the experimenter-defined Class 3 in Pre-SRT-WoR did so in Pre-SRT-WR.

TBR-WR in Both Sequences

The participants used from 150 to 840 trials (see Table 2). The participants used from 150 trials to 840 to reach the criterion of 100% correct trials in the block with 0% programmed consequences in Sequence 1 and in Sequence 2 they used from 150 trials to 450. The mean of the trials used varied from 309 in Sequence 1 to 243 in Sequence 2. Independent-samples t-tests were conducted to compare the two sequences with respect to numbers of baseline training trials used before the thinning of consequences (establishment) and the numbers of baseline training trials used during the phase with the thinning (maintenance) of consequences in the H-MTS training. There was no significant difference in establishing conditional discrimination between Sequence 1 (M = 231,00, SD = 205,45) and Sequence 2 (M = 183,00, SD = 93,22); $t(18) = 0.673$, $p = .510$. Furthermore, there were no significant differences in the maintenance of conditional discrimination between Sequence 1 (M = 78, SD = 37,95) and Sequence 2 (M = 60, SD = 0,00); $t(18)$ = 1,500, $p = .151$.

Post-Class Formation Sorting Tests in Sequence 1 (1^{st} Block)

SRT-WoR. Two out of the ten participants (15610 and 15623) sorted all three experimenter-defined 3-member classes, and the other participants marked the stimuli in two to four groups. Participant 15601 sorted Class 2, and Participant 15622 sorted Class 3. A chi-square test revealed no significant results with respect to mastering all experimenter-defined classes between Pre-SRT-WoR to SRT-WoR.

SRT-WR. . Seven of the ten participants sorted all the experimenter-defined three-member classes. The participants who did not sort all three experimenter-defined classes sorted the stimuli into two or three groups. Two participants (15618 and 15608) sorted the experimenter-defined Class 1 (see Table 2). A chi-square test revealed significant (p = .001) results from Pre-SRT-WR to SRT-WR with respect to mastering all experimenter-defined classes.

Post-Class Formation Sorting Tests in Sequence 1 (2nd Block)

Post-SRT-WoR. Three of the ten participants sorted all three experimenter-defined 3-member classes. The participants who did not sort all the experimenter-defined classes sorted the stimuli into four groups. Participant 15601 sorted one class as the experimenter-defined class 2, and Participant 15607 sorted the experimenter-defined Class 3.

Post-SRT-WR. Seven of the ten participants sorted all three experimenter-defined 3-member classes. The Participants who did not sort all the experimenter-defined classes sorted the stimuli in two or three groups. Participant 15618 sorted the experimenter-defined Class 1.

Consistent Responding in Post-Class Formation Sorting Tests in Sequence 1

Table 4 shows that Participant 15618 repeated one marked group from Pre-SRT-WoR to SRT-WoR, namely, the stimuli B1, B2. Participant 15601 repeated two marked groups form Pre-SRT-WoR to SRT-WoR, (i.e., A1 and B1 in one group and A3 alone) and also repeated the sorting from SRT-WoR to Post-SRT-WoR. Participant 15611 repeated the marked group of A1, A2, A3, B3, C3 from Pre-SRT-WR to SRT-WoR and repeated the exact same sorting from SRT-WoR to Post-SRT-WoR. Participant 15617 repeated the marked group of C1, C3 from Pre-SRT-WR to SRT-WoR, and Participant 15608 made the exact same sorting in Post-SRT-WR as in Post-SRT- WoR.

Class Formation in H-MTS-WR-Test

Eight of the ten participants responded in accordance with stimulus equivalence during the MTS-test and reached the criteria of 95% correct trials. Participants 15618 and 15608 did not respond in accordance with stimulus equivalence, but Participant 15617 now responded in accordance with stimulus equivalence, though this participant did not sort correctly in the prior sorting test.

Analysis of participant-defined responding in H-MTS-WR. In Table 3, the incorrect trials are analyzed trial by trial to see if they are due to consistent participant-defined responding. In the upper-left matrices, Participant 15608's choices for all trial types are shown. In the baseline trials, when sample A3 was presented, the participant chose C1 twice and C3 once. In the symmetry trials, when sample C3 was presented, the participant chose A2 twice and A3 once. The rest of the choices in the baseline and symmetry trials were in accordance with the experimenter-defined classes. In the transitivity and equivalence trials, when the sample was A3, the participant chose B1 in all three trials, but when the sample presented was B1, the participant chose A3 only once and A1 twice. When the sample was B3, the participant chose A2 twice and A3 once. However, when the sample was A2, the participant chose stimuli in accordance with the experimenter-defined classes. These results show no consistent participant-defined classes for participant 15608. We analyzed the results for participant 15618 in the same way, as can be seen in the Table 3 upper-right matrices, and found no participant-defined classes here either.

Results of Post-Class Formation Sorting tests in Sequence 2

Post-SRT-WoR. One of the ten participants (15624) formed all the experimenter-defined classes, while the others marked from three to five groups of stimuli not corresponding with the experimenter-defined classes. Participant 15612 marked the groups of stimuli corresponding to the experimenter-defined Class 2. The chi-square tests revealed no significant differences between the results in Pre-SRT-WoR and Post-SRT-WoR.

Post-SRT-WR. Seven of the ten participants marked the groups of stimuli corresponding to all the experimenter-defined classes, and the others sorted the stimuli into two to five groups not corresponding to the experimenter-defined classes. A chi-square test was conducted to compare the results on sorting tests between the Pre-SRT-WR and Post-SRT-WR and it was found to be statistically significant (p = .025)

Consistent Responding Between Sorting Tests in Sequence 2

Some participants who did not sort in accord with the experimenter-defined classes had some consistent responding. In Post-SRT-WoR, Participant 15621 marked one group equal to one in Pre-SRT-WR, and Participant 15616 marked two equal groups of stimuli in Pre-SRT-WR and Post-SRT-WoR. Participant 15615 marked all four groups as in Pre-SRT-WoR and Post-SRT-WoR (see Table 4). Participant 15609 marked the stimuli in the exact same groups as in Post-SRT-WoR. Participant 15603 marked two groups that differed from any of the participants' prior classes.

Results of Class Formation in Sequence 2

In the first class formation test **(**H-MTS-WR-test-1) seven of the ten participants responded in accordance with stimulus equivalence in H-MTS-WR-test-1, but Participants 15624, 15603 and 15609 did not. In the second class formation test (H-MTS-WR-test-2) eight of the ten participants responded in accordance with stimulus equivalence during H-MTS-WR-test-2, but Participants 15603 and 15609 did not. No consistent participant-defined responding was found for any of the participants or in any of the tests when we analyze the results trial by trial (see Table 3).

Class Formation Between Test Types and Sequences

In Figure 4, the formation of classes is depicted. In the right column, the diagrams show the different combinations of class formations done in Sequence 1 and how many participants did each combination. In the left column is shown the different class formations created in Sequence 2. By counting the number of single classes depicted by the bars, we see a formation of 8 of 30 possible classes in SRT-WoR and a formation of 11 of 30 possible classes in Post-SRT-WoR among all the participants in Sequence 1, whereas we see a formation of 4 of 30 possible classes in Post-SRT-WoR among all the participants in Sequence 2.

Reaction Time

For analyzing the reaction time within each H-MTS test, independent samples t-tests were performed. The analysis included calculations of the differences between the medians of the first five and the last five trials in each trial type (i.e., baseline, symmetry, and transitivity/equivalence) in the H-MTS tests. The t-tests were conducted for all three H-MTS tests, and no significant differences between the first five and the last five trials were found in any of the trial types in any of the H-MTS tests.

To analyze the differences between the three H-MTS tests, an analysis of variance (ANOVA) was conducted between all three H-MTS tests with respect to the medians of the reaction times for the first five trials and for the last five trials for each trial type. The ANOVA showed that there was no significant difference in reaction time when comparing the different trial types between the two sequences. Based on the t-tests and the ANOVA analyzes the results on the reaction time for all 20 participants were combined.

Figure 3 shows the mean median reaction time for all twenty participants, showing the last five training trials of the training phase and the first five (in black) and last five (in gray) baseline trials testing for the directly trained relations, the first five and the last five trials testing for symmetry and the first five and last five testing for transitivity or global equivalence. The last bar shows the mean medians for all types of errors made in the test. By a visual inspection of Figure 3, we see that the reaction time increases from the baseline to symmetry trials and from the symmetry to equivalence/ transitivity trials. We see a decrease in the reaction time from the first five trials to the last five in all trial types, and we see that the reaction time for incorrect choices is higher than for all types of correct trials.

Interobserver Agreement

Two independent observers scored all the participants' sorting of the stimuli. 99 of 100 sorting tests were scored equal, making an interobserver agreement (IOA) of 99%. The observers scored the Pre-SRT-WR of Participant 15609 differently. The observers agreed on three separate marker groupings in which six of the stimuli were placed, but disagreed about the last three stimuli if they formed one or two groups. The observers agreed that the sorting as a total was far from close to any of the experimenter-defined classes. Therefore, this disagreement is no threat to the validity.

Discussion

Sorting of stimuli WoR and Immediate Emergence of Stimulus Classes

To address attending behavior in the present experiment, one question asked was if H-MTS training with a set of stimuli (i.e., Chinese characters) in which the stimuli in the same classes had identical features (i.e., radicals) would have any impact on the immediate emergence of class formation in the same set of stimuli but without the identical features. This was evaluated by the use of sorting tests and by comparing the immediate emergence of class formation between Sequence 1 and Sequence 2.

The sorting tests gave the following results. First, we found that two of the 10 participants demonstrated the immediate emergence of all the experimenter-defined classes of stimuli in the SRT-WoR test in Sequence 1, while one of the 10 participants showed all the experimenter-defined classes in Pre-SRT-WoR in Sequence 2. Second, we found that the difference in results from Pre-SRT-WoR to SRT-WoR in sequence 1 and from Pre-SRT-WoR to Post-SRT-WoR in sequence 2 were not significant in any of the sequences. Third, we found the immediate emergence of eight of the 30 possible single experimenter-defined classes in Sequence 1 when SRT-WoR was conducted right after TBR-WR. We also found that 11 of the 30 possible single experimenter-defined classes emerged in Sequence 1 in the Post-SRT-WoR after the H-MTS-WR-test, whereas in Sequence 2, when the Post-SRT-WoR was conducted after the H-MTS-WR-test-1 only four of the 30 possible single experimenter-defined classes emerged.

These findings can be interpreted in different ways. It is possible that the lower results on sorting in Sequence 2 are related to the fact that the sorting is conducted after the H-MTS-WR-test-1. By adding the H-MTS-WR-test-1 right after the TBR-WR in Sequence 2, the participants are in the H-MTS-WR-test-1 exposed to the radicals an increased number of times. Thereby, what seems to be a slightly less attending behavior in Sequence 2 compared to in Sequence 1 could be due to blocking, as argued by Donahoe and Palmer (1994).

Another reason for the attending behavior to occur to a lower degree could be if the participants failed to sense that the stimuli present in the SRT-WoR were actually partly the same as those used in TBR-WR. This could also be a reason for the low results in SRT-WoR in both sequences.

To investigate how much or if the participants are focusing on or observing the arbitrary part of the stimuli under the TBR-WR, it would be interesting to use eye-tracking equipment. Investigations in attending behavior, MTS and eye-tracking have lately been summarized in the following way by Hansen and Arntzen (2015).

> Furthermore, we point to evidence which suggests that attending, looking, observing, and perceiving operate on something of a functional continuum, with attending and looking - or vice versa - at one end of the scale, with differentially reinforced ocular observing responses further along, and with perceiving at the other end of the continuum. Thus, attending constitutes a controlling relationship between the visual contact that meets the eye and a visual discriminative stimulus, established and maintained by conditioned reinforcement (p. 244).

This taken into consideration together with the results of this experiment, it would be interesting to replicate the study using eye-tracking to examine how the identical radicals in the compound stimuli affect the observing behavior.

Sorting of stimuli WR and Immediate Emergence of Stimulus Classes

In Sequence 1, seven of the ten participants showed all the experimenter-defined classes of the stimuli WR when the sorting tests WR were conducted after TBR-WR. Therefore, in Sequence 1, the sorting tests of stimuli WR showed the immediate emergence of all the three experimenter-defined classes for 70% of the participants. In Sequence 2, the sorting tests were administered after both the TBR-WR and the H-MTS-WR-test-1 were conducted. Seven of the 10 participants showed the immediate emergence of all three experimenter-defined classes of stimuli WR in the H-MTS-WR-test-1, and the same number showed the emergence of the three experimenter-defined classes of stimuli WR in the subsequent sorting test. Across the groups, sorting tests WR showed the emergence of all three stimuli classes by 14 of the 20 participants. Compared to Pre-SRT-WR, the results showed high and significant results for sorting the stimuli WR after only TBR-WR as well as after conducting both TBR-WR and H-MTS-WR-test-1. These results replicated the prior findings of Arntzen et al. (2015) and Arntzen et al. (2017) showing that a sorting test can track class formation by sampling a subset of emergent relations. However, when comparing the yields with those of prior experiments, we should take into consideration that this was a hybrid of matching-to-sample with stimuli having some identical features in each class. Additionally, two participants in Pre-SRT-WR marked all the groups of the three experimenter-defined classes. These participants were not excluded from the experiment because the main purpose of this experiment was to examine the sorting of the stimuli WoR after the participants had learned the relations of the stimuli WR. For the same reason, the MTO training structure was used in this study to give high yields, whereas Arntzen et al. (2015) and Arntzen et al. (2017) used an LS training structure.

The reaction time data displayed in Figure 3 shows that across the two sequences, there was an increase in the mean of median reaction time from the last five training trials to the first five trials of testing for the trained relations. The mean of median reaction time for the trials testing for symmetry was higher for the first five trials than for the first five trials testing the trained relations, and an increase was also shown from the first five trials of symmetry to the first five of the equivalence/transitivity trials. A decrease in the mean of medians was also observed in all three relations from the first five trials to the last five trials tested. These data replicate those of other studies on the reaction time in trials testing for stimulus equivalence relations (e.g., Arntzen et al., 2011; Arntzen et al., 2007; Arntzen et al., 2010; Arntzen & Hansen, 2011; Eilifsen & Arntzen, 2009; Spencer & Chase, 1996). These findings are interesting taking into consideration that we used an H-MTS procedure with some identical features in the compound stimuli. One assumption should be that the reaction time was equal under the whole test if the identical parts of the compound stimuli controlled the comparison choice. However, when the reaction time pattern in the present study resembled patterns from studies testing equivalence relations (e.g., Arntzen et al., 2011; Arntzen et al., 2007; Arntzen et al., 2010; Arntzen & Hansen, 2011; Eilifsen & Arntzen, 2009; Spencer & Chase, 1996), it suggests that even though the compound stimuli had some identical parts, the identical parts of the stimuli did not fully control the behavior. Rather, the stimuli could be assumed as abstract stimuli, and thereby they constituted an arbitrary matching task and equivalence class formation.

To determine when and if the mean of the median reaction time would level out between the different trial types with respect to identical features in the compound stimuli, it would be interesting to replicate the study and measure the reaction time while we gradually increase the percentage of identical parts of the compound stimuli.

Equivalence Classes, Stimulus Classes, and Sorting WR

Seven participants in Sequence 1 and under SRT-WR showed the immediate class formation of all three classes, and the same participants also responded correctly according to the experimenter-defined classes in a subsequently administered H-MTS-WR-test (see Table 2 and Figure 4). This shows what would be called equivalence classes if we had used stimuli with no physical similarities. However, as the reaction time in the H-MTS tests showed the same pattern as for arbitrary matching, we find it appropriate to interpret the results as if the task in the H-MTS-WR-test was arbitrary matching. If this interpretation is correct, the stimulus classes documented by the sorting tests seem to be predictive of the equivalence classes that were documented in the H-MTS-WR-test.

Consistent Responding in Sorting Tests

Some participants who showed the absence of experimenter-defined classes anyway sorted the classes with some consistency. When we compared the marked groups between the different sorting tests WR (see Table 4), we found that Participant 15618 repeated A3, C3 from SRT-WoR to SRT-WR in Sequence 1 and Participant 15608 repeated all the groups form Post-SRT-WoR to Post-SRT-WR. Similarly, in sequence 2, Participant 15615 and Participant 15609 repeated the all the sorted groups from Post-SRT-WoR to Post-SRT-WR. These findings indicate a consistent class formation defined by the participant. However, when we analyzed the H-MTS tests trial by trial in Table 3, we did not find the same participant-defined classes.

Delayed Emergence

Figure 4 shows that Participant 15617 and Participant 15624 showed delayed emergence of classes. For Participant 15617, in Sequence 1 in the SRT-WR after TBR-WR, no experimenter-defined classes emerged, but they did so in the follow-up H-MTS-WR-test, and in the Post-SRT-WR, all three experimenter-defined classes emerged. For Participant 15624 in Sequence 2, a delay was seen in H-MTS-WR-test-1; after TBR-WR, no experimenter-defined classes had emerged, but they all emerged in the following tests. Sidman (1994) wrote,

> " … delayed emergence might reflect the fact that stimuli can belong to other classes in addition to the experimentally established equivalence classes that are being tested. (pp. 274–279). Some of these classes may be products of the subject's extraexperimental history (Sidman, 1992a, pp. 23–24); others, although irrelevant to the aims of the study, may nevertheless have been created within it". (p. 511)

Nevertheless, when we compared the outcomes for delayed emergence in the two sequences when either sorting tests or H-MTS-tests are presented right after TBR-WR, we found no difference in the sensitivity for delayed emergence between the two sequences.

Pre-Sorting

In Pre-SRT-WoR, we found that three participants sorted the experimenter Class 3, but we suggest this was by chance. If there are some common features of the stimuli in Class 3, we would expect the participants to react to the same features in Pre-SRT-WR, but none of the participants did so.

In Table 2 and in Figure 4, we see that Participant 15620 documented the stimulus classes in the sorting test and in the follow-up H-MTS-WR-test but did not document the classes again in the follow-up sorting test Post-SRT-WR. An interpretation can be that the sorting performed in Pre-SRT-WoR interfered with the newly achieved class-based since the participant marked the exact same groups as in Pre-SRT-WoR (see Table 4).

Participants 15618 and 15601 both repeated marked groups from Pre-SRT-WoR to SRT-WoR, Participants 15611 and 15617 repeated marked groups from Pre-SRT-WR, and Participant 15609 repeated two groups of stimuli from Pre-SRT-WoR to Post-SRT-WoR in Sequence 2. These results suggest that the pre-sorting tests can affect the overall outcome and thus should be used with caution.

Verbal Reports From Debriefing

Participant 15610 marked all the nine stimuli in one group in both Pre-SRT-WoR and

Pre-SRT-WR. During the debriefing, the participant said that this was due to having no idea of how to sort the symbols.

Participant 15615 marked all the same groups in all the sorting tests but one: between Pre-SRT-WoR and Pre-SRT-WR, only one stimulus was placed differently. At the same time, the participant did both the H-MTS-WR-test-1 and H-MTS-WR-test-2 in accordance with the experimenter-defined classes. During the debriefing, the participant reported that this was done intentionally, convinced that the remembrance of the first sorting was the main purpose of all the sorting tasks. The participant explained that she thought this because memory was mentioned in the information sheet. Although the information sheet was not meant to be an instruction, the verbal report from Participant 15615 indicates that it had a function like this for this participant. We do not know if it had this function for any other of the participants, but no one but Participant 15615 exhibited consistent results of this type. Sidman (1994) writes that instructions to the subject "may establish a context that brings into play historical contingencies that interact with or completely override current experimental contingencies" (p. 510). Clearly, this makes it even more pertinent to be careful with instructions and information given to participants before an experiment.

Limitations

There are some limitations to the present study. We had only 10 participants in each group, and we used statistical calculations. In further research, we should increase the number of participants in each group to reduce the possibility of coincidences that might influence the data. Another limitation could be the information sheet, which influenced at least one participant to repeat the results of the first sorting tests throughout the experiment. However, we did not find similarly consistent results in any other participant's performance.

Further Research

This experiment opens the opportunity for further research to be proposed: These include to investigate with eye-tracking equipment how much or if the participants are focusing on the arbitrary parts of the stimuli under the TBR-WR and reveal if there is a correlation to the outcome on the stimulus equivalence test; to employ an MTS-WoR after conducting the TBR-WR and sorting WoR have been done to investigate if the classes of stimuli WoR that emerged under sorting have the properties of equivalence classes; and to investigate how large a part of the stimuli that needs to be identical before the results from the MTS test would resemble an identical matching test with an assumed equal reaction time.

Summary

This experiment showed the immediate emergence of stimulus classes after TBR-WR in 70% of the participants for both test types (i.e., sorting tests and MTS tests). However, the immediate emergence of all stimulus classes WoR was only seen in 20% of the participants in Sequence 1 and 10% in Sequence 2. With respect to the single experimenter-defined classes, we found higher yields for sorting tests WoR in Sequence 1 than in Sequence 2. Thus, we found higher yields when sorting WoR was conducted right after TBR-WR.

That we used stimuli with some identical features should be taken into consideration when comparing this study with prior studies and considering equivalence classes. However, with the results on reaction time showing that the pattern in the mean of medians in this study resembled those of patterns from studies testing equivalence relations with arbitrary stimuli, we found it reasonable to compare this study to other studies using conditional discrimination with arbitrary stimuli. Thus, the findings in this experiment extend the findings of Arntzen et al. (2015) and Arntzen et al. (2017) and increase the evidence for the reliability of documenting stimulus class formation with sorting tests by showing that the immediate emergence of stimulus classes WR can be tracked with sorting tests. When we employed an H-MTS-WR-test after the sorting test WR in Sequence 1, it documented (what can be assumed to be) equivalence classes for the classes previously documented by sorting. Therefore, it is possible that the classes shown by the sorting test also had the properties of equivalence classes.

The experiment also showed that the immediate emergence of stimulus classes with identical features removed before the sorting test could be tracked with sorting tests; however, we do not know if these classes have the properties of equivalence classes.

References

Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: methodological issues. *European Journal of Behavior Analysis, 13*, 123–135. Retrieved from http://www.ejoba.org/.

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. doi:10.1080/15021149.2011.11434398

Arntzen, E., Galaen, T., & Halvorsen, L. R. (2007). Different retention intervals in delayed matching-to-sample: Effects of responding in accord with equivaence. *European Journal of Behavior Analysis, 8*, 177–191. doi:10.1080/15021149.2007.11434281

Arntzen, E., Granmo, S., & Fields, L. (2017). The relation between sorting tests and matching-to-sample tests in the formation of equivalence classes. *The Psychological Record*. doi:10.1007/s40732-016-0209-9

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://thepsychologicalrecord.siuc.edu/index.html.

Arntzen, E., & Hansen, S. (2011). Training structures in the formation of equivalence classes. *European Journal of Behavior Analysis, 12*, 483–503. doi:10.1080/15021149.2011.11434397

Arntzen, E., Norbom, A., & Fields, L. (2015). Sorting: An alternative measure of class formation? *The Psychological Record, 65*, 615–625. doi:10.1007/s40732-015-0132-5

Augustson, E. M., Dougher, M. J., & Markham, M. R. (2000). Emergence of conditional stimulus relations and transfer of respondent eliciting functions among compound stimuli. *The Psychological Record, 50*, 745–770. Retrieved from ISI:000165227100008.

Bentall, R. P., Jones, R. M., & Dickins, D. W. (1998). Errors and response latencies as a function of nodal distance in 5-member eguivalence classes. *The Psychological Record, 49*, 93–115.

Catania, A. C. (2013). *Learning* (5th ed.). Cornwall-on-Hudson, NY: Sloan Pub.

Debert, P., Matos, M. A., & McIlvane, W. (2007). Conditoional relations with compound abstract stimuli using a go/no-go procedure. *Journal of the Experimantal Analysis of Behavior, 87*, 89–96. doi:10.1901/jeab.2007.46-05

Donahoe, J. W., & Palmer, D. C. (1994). *Learning and complex behavior* (V. P. Dorsel Ed.). Boston: Allyn and Bacon.

Dymond, E., & Rehfeldt, R. A. (2001). Supplemental measures and derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin, 19*, 8–12. Retrieved from http://www.eahb.org/11/wp-content/uploads/2014/2003/Dymond-Rehfeldt.pdf.

Eilifsen, C., & Arntzen, E. (2009). On the role of trial types in tests for sitmulus equivalence. *European Journal of Behavior Analysis, 10*, 187–202. Retrived from http://www.ejoba.org.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*(1), 31–60. doi:10.1037/0033-2909.134.1.31

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: Plenum Press.

Hansen, S., & Arntzen, E. (2015). Fixating, attending, and observing: a behavior analytic eye-movement analysis. *European Journal of Behavior Analysis, 16*(2), 229–247. doi:10.1080/15021149.2015.1090750

Hull, C. L. (1920). *Quantitative aspects of the evolution of concepts : an experimental study.* (Ph D), University of Wisconsin, Princeton, N.J. ; Lancaster, Pa.

Mcleod, P. (1977). A dual task response modality effect: Support for multiprocessor models of attention. *Quarterly journal of experimental psychology, 29*(4), 651–667. doi:10.1080/14640747708400639

Recorla, R. A., & Wagner, A. R. (1972). A Theory of Pavlovian Conditioning: Variations in the Effevtiveness of Reinforcement and Nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current Research and Theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Reynolds, G. S. (1961). Attention in the pigeon. *J Exp Anal Behav, 4*, 203-208. doi:10.1901/jeab.1961.4-203

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

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

Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.

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

Table 1

Sequences and Conditions

Sequence 1:

Pre-SRT-WoR – Pre-SRT-WR – TBR-WR – SRT-WoR – SRT-WR – H-MTS-WR-test – Post-SRT-WoR – Post-SRT-WR

Sequence 2:

Pre-SRT-WoR – Pre-SRT-WR – TBR-WR – H-MTS-WR-test-1– Post-SRT-WoR – Post-SRT-WR – H-MTS-WR-test-2

Note. The participants were exposed to different conditions depending on one of two sequences. The order of sequences is shown from left to the right. Pre-SRT = pre-class formation sorting test; TBR = training of baseline relations; SRT = sorting test after training of baseline relations; MTS = Matching-To-Sample test; Post-SRT = sorting test after the MTS-test; WoR = without radicals; WR = with radicals; H = hybrid (see text for explanation); test-1 = first test; test-2 = second test.

Table 2

Results From Sequence 1 and Sequence 2

Sequence 1

PN	Pre-SRT-WoR	Pre-SRT-WR	TBR-WR	SRT-WoR	SRT-WR	H-MTS-WR-test	Post-SRT-WoR	Post-SRT-WR
15610	333	333	390	**300 030 003**	**300 030 003**	Y	**300 030 003**	**300 030 003**
15623	210 120 **003**	111 111 111	330	**300 030 003**	**300 030 003**	Y	**300 030 003**	**300 030 003**
15622	310 020 **003**	**300** 020 011 002	210	210 120 **003**	**300 030 003**	Y	**300 030 003**	**300 030 003**
15601	200 112 020 001	201 120 010 002	150	200 **030** 102 001	**300 030 003**	Y	200 **030** 102 001	**300 030 003**
15607	202 131	203 130	330	230 103	**300 030 003**	Y	210 120 **003**	**300 030 003**
15611	110 101 102 020	110 110 113	180	220 113	**300 030 003**	Y	220 113	**300 030 003**
15617	202 131	111 121 101	150	101 120 112	331 002	Y	111 121 101	**300 030 003**
15620	201 111 021	111 120 102	840	210 111 012	**300 030 003**	Y	201 120 012	111 120 102
15618	211 110 012	200 **030** 101 002	240	111 110 110 002	**300** 031 002	N	111 121 101	**300** 031 002
15608	211 122	222 111	270	203 130	**300** 033	N	301 032	301 032

Sequence 2

PN	Pre-SRT-WoR	Pre-SRT-WR	TBR-WR	H-MTS-WR-test-1	Post-SRT-WoR	Post-SRT-WR	H-MTS-WR-test-2
15612	111 120 102	**300 030 003**	210	Y	201 **030** 102	**300 030 003**	Y
15619	201 120 012	**300 030 003**	210	Y	110 120 103	**300 030 003**	Y
15621	111 120 102	**300 030** 002 001	150	Y	110 120 101 002	**300 030 003**	Y
15616	200 110 011 012	101 110 120 002	210	Y	100 110 101 020 002	**300 030 003**	Y
15605	230 103	203 130	240	Y	211 120 002	**300 030 003**	Y
15614	210 121 002	211 122	360	Y	210 101 022	**300 030 003**	Y
15615	200 110 021 002	100 110 121 002	150	Y	200 110 021 002	200 110 021 002	Y
15624	311 022	310 020 **003**	210	N	**300 030 003**	**300 030 003**	Y
15603	120 210 **003**	211 122	450	N	101 110 120 002	230 103	N
15609	110 110 101 002 010	200 101 020 012	240	N	200 101 020 002 010	200 101 020 002 010	N

Note. The top section shows the results of Sequence 1, and the bottom section shows the results of Sequence 2. The first column indicates the participant numbers (PN). Each row shows the results for a given participant in all the phases of the experiment. The different phases of the experiment are named in the headings of the columns. Under the headlines containing SRT (sorting tests), the 3-digit figures indicate the clusters of the stimuli the participant produced. Reading from left to right, the 1st digit shows how many stimuli there were to be found in the cluster from Class 1, the 2nd digit from Class 2 and the 3rd from Class 3. The digits in **bold** font show that the classes matched the experimenter-defined classes. TBR- = training of baseline relations showing how many trials the participant used to reach the criterion; MTS-test- = Equivalence class formation shown in the Matching-To-Sample test and indicated by Y for those who formed equivalence classes and N for those who did not.

Table 3

Trial-by-Trial Analysis of Emergent Relations

15608	Test type	Comparisons					
	Baseline	C1	C2	C3	C1	C2	C3
Samples	A1	3					
	A2		3				
	A3	2		1			
	B1				3		
	B2					3	
	B3						3
	Test type	Comparisons					
	Symmetry	A1	A2	A3	B1	B2	B3
Samples	C1	3					
	C2		3				
	C3		2	1			
	C1				3		
	C2					3	
	C3						3
	Test type	Comparisons					
	Equivalence	B1	B2	B3	A1	A2	A3
Samples	A1	3					
	A2		3				
	A3	3					
	B1				2		1
	B2					3	
	B3					2	1

15618	Test type	Comparisons					
	Baseline	C1	C2	C3	C1	C2	C3
Samples	A1	3					
	A2		3				
	A3			3			
	B1				3		
	B2					3	
	B3						3
	Test type	Comparisons					
	Symmetry	A1	A2	A3	B1	B2	B3
Samples	C1	3					
	C2		3				
	C3			3			
	C1				3		
	C2					3	
	C3						3
	Test type	Comparisons					
	Equivalence	B1	B2	B3	A1	A2	A3
Samples	A1	3					
	A2		3				
	A3			3			
	B1				3		
	B2					3	
	B3				1	2	0

15603 1.test	Test type	Comparisons					
	Baseline	C1	C2	C3	C1	C2	C3
Samples	A1	3					
	A2		2	1			
	A3			3			
	B1				3		
	B2					3	
	B3						3
	Test type	Comparisons					
15603	Symmetry	A1	A2	A3	B1	B2	B3
Samples	C1	3					
	C2		3				
	C3			3			
	C1				3		
	C2					3	
	C3				1		2
	Test type	Comparisons					
	Equivalence	B1	B2	B3	A1	A2	A3
Samples	A1	3					
	A2		3				
	A3			3			
	B1				1		2
	B2					3	
	B3						3

15603 2.test	Test type	Comparisons					
	Baseline	C1	C2	C3	C1	C2	C3
Samples	A1	1					
	A2		2	1			
	A3			3			
	B1				3		
	B2					3	
	B3						3
	Test type	Comparisons					
	Symmetry	A1	A2	A3	B1	B2	B3
Samples	C1	2		1			
	C2		3				
	C3			3			
	C1				3		
	C2					3	
	C3						3
	Test type	Comparisons					
	Equivalence	B1	B2	B3	A1	A2	A3
Samples	A1	3					
	A2		3				
	A3			3			
	B1				1	1	1
	B2					3	
	B3						3

15609 1.test	Test type	Comparisons					
	Baseline	C1	C2	C3	C1	C2	C3
Samples	A1	3					
	A2		3				
	A3			3			
	B1				3		
	B2					3	
	B3						3
	Test type	Comparisons					
15609	Symmetry	A1	A2	A3	B1	B2	B3
Samples	C1	1		2			
	C2		3				
	C3			3			
	C1				2		1
	C2					3	
	C3						3
	Test type	Comparisons					
	Equivalence	B1	B2	B3	A1	A2	A3
Samples	A1	2		1			
	A2		3				
	A3			3			
	B1				3		
	B2					3	
	B3				1		2

15609 2.test	Test type	Comparisons					
	Baseline	C1	C2	C3	C1	C2	C3
Samples	A1	3					
	A2		3				
	A3	1		2			
	B1				3		
	B2					3	
	B3						3
	Test type	Comparisons					
	Symmetry	A1	A2	A3	B1	B2	B3
Samples	C1	1		2			
	C2		3				
	C3			3			
	C1				3		
	C2					3	
	C3						3
	Test type	Comparisons					
	Equivalence	B1	B2	B3	A1	A2	A3
Samples	A1	2		1			
	A2		3				
	A3			3			
	B1				3		
	B2					3	
	B3						3

15624 1.test	Test type	Comparisons					
	Baseline	C1	C2	C3	C1	C2	C3
Samples	A1	3					
	A2		3				
	A3			3			
	B1				3		
	B2					3	
	B3						3
	Test type	Comparisons					
	Symmetry	A1	A2	A3	B1	B2	B3
Samples	C1	3					
	C2		3				
	C3			3			
	C1				3		
	C2					3	
	C3					1	2
	Test type	Comparisons					
	Equivalence	B1	B2	B3	A1	A2	A3
Samples	A1	3					
	A2		3				
	A3	2		1			
	B1				3		
	B2					3	
	B3						3

Note. This table shows matrices for all trial types; baseline trials in the upper, symmetry in the middle and transitivity/equivalence in the lower panel of each matrix. One matrix is shown for each of the five participants (participant number in the upper left corner) who did not reach the criterion for responding in accordance with equivalence in each H-MTS-test. The matrices show the frequencies of selecting each comparison stimulus (columns) in the presence of each sample stimulus (rows). Each trial type was presented three times, so the presence of the number "3" in the marked squares between the sample and comparison shows that the relation is in accordance with the experimenter-defined classes. The numbers outside the squares show participant-defined relations.

Table 4

Notations of Stimuli in the Marked Groups of Stimuli

Sequence 1

PN	Pre-SRT-WoR	Pre-SRT-WR	SRT-WoR	SRT-WR	Post-SRT-WoR	Post-SRT-WR
15610	A1 A2 A3 B1 B2 B3 C1 C2 C3	A1 A2 A3 B1 B2 B3 C1 C2 C3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**	**A1 B1 C1, A2 B2 C2, A3 B3 C3**	**A1 B1 C1, A2 B2 C2, A3 B3 C3**	**A1 B1 C1, A2 B2 C2, A3 B3 C3**
15623	A1 B1 C2, C1 A2 B2, **A3 B3 C3**	A1 A2 A3, B1 C2 C3, C1 B2 B3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**	**A1 B1 C1, A2 B2 C2, A3 B3 C3**	**A1 B1 C1, A2 B2 C2, A3 B3 C3**	**A1 B1 C1, A2 B2 C2, A3 B3 C3**
15622	A1 B1 C1 B2, A2,C2 **A3 B3 C3**	**A1 B1 C1,** A2 C2, B2 C3, A3 B3	B1 C1 C2, A1 A2 B2, **A3 B3 C3**	**A1 B1 C1, A2 B2 C2, A3 B3 C3**	**A1 B1 C1, A2 B2 C2, A3 B3 C3**	**A1 B1 C1, A2 B2 C2, A3 B3 C3**
15601	A1 B1, C1 B2 B3 C3, A2 C2, A3	A1 B1 A3, C1 A2 B2, C2, B3 C3	A1 B1, **A2 B2 C2,** C1 B3 C3, A3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**	A1 B1, **A2 B2 C2,** C1 B3 C3, A3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**
15607	A1 C1 A3 B3, B1 A2 B2 C2 C3	A1 C1 A3 B3 C3, B1 A2 B2 C2	A1 B1 A2 B2 C2, C1 A3 B3 C3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**	A1 C1 C2, B1 A2 B2, **A3 B3 C3**	**A1 B1 C1, A2 B2 C2, A3 B3 C3**
15611	A1 B2, B1 A3, C1 B3 C3, A2 C2	B1 B2, C1 C2, A1 A2 A3 B3 C3	B1 B2 C1 C2, A1 A2 A3 B3 C3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**	B1 B2 C1 C2, A1 A2 A3 B3 C3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**
15617	A1 C1 A3 C3, B1 A2 B2 C2 B3	A1 B2 A3, B1 A2 C2 B3, C1 C3	C1 C3, A1 A2 B2, B1 C2 A3 B3	A1 B1 C1 A2 B2 C2 A3, B3 C3	B1 C2 B3, A1 A2 A3 B2, C1 C3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**
15620	A1 C1 B3, B1 A2 C3, B2 C2 A3	A3 B1 B2, A2 C1 C2, A1 B3 C3	B1 C1 A2, A1 B2 C3, C2 A3 B3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**	A1 B1 C3, C1 A2 B2, C2 A3 B3	A1 A2 A3, B1 B2 C2, C1 B3 C3
15618	A1 C1 C2 A3, B1 B2, A2 B3 C3	A1 B1, **A2 B2 C2,** C1 C3, A3 B3	A1 A2 B3, B1 B2, C1 C2, A3 C3	**A1 B1 C1,** A2 B2 C2 B3, A3 C3	B1 C2 C3, A1 A2 B2 B3, C1 A3	**A1 B1 C1,** A2 B2 C2 A3, B3 C3
15608	A1 B1 B2 B3, C1 A2 C2 A3 C3	A1 B1 B2 C2 B3 C3, C1 A2 A3	A1 C1 A3 B3 C3, B1 A2 B2 C2	**A1 B1 C1,** A2 B2 C2 A3 B3 C3	A1 B1 C1 A3, A2 B2 C2 B3 C3	A1 B1 C1 A3 A2 B2 C2 B3 C3

Sequence 2

PN	Pre-SRT-WoR	Pre-SRT-WR	Post-SRT-WoR	Post-SRT-WR
15612	A1 B2 C3, B1 A2 C2, C1 A3 B3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**	B1 C1 A3, **A2 B2 C2,** A1 B3 C3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**
15619	A1 B1 C3, C1 A2 C2, B2 A3 B3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**	C1 A2, A1 B2 C2, B1 A3 B3 C3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**
15621	B1 C2 B3, A1 A2 B2, C1 A3 C3	**A1 B1 C1, A2 B2 C2,** A3 B3, C3	B1 C2, A1 A2 B2, C1 C3, A3 B3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**
15616	A1 C1, B1 C2, B2 C3, A2 A3 B3	C1 C3, A1 B2, B1 A2 C2, A3 B3	B1, A1 B2, C1 C3, A2 C2, A3 B3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**
15605	A1 B1 A2 B2 C2, C1 A3 B3 C3	A1 B1 A3 B3 C3, C1 A2 B2 C2	A1 C1 B2 C3, B1 A2 C2, A3 B3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**
15614	B1 C1 C2, A1 A2 B2 C3, A3 B3	B1 C1 C2 A3 A1 A2 B2 B3 C3	A1 C1 C2, B1 A3, A2 B2 B3 C3	**A1 B1 C1, A2 B2 C2, A3 B3 C3**
15615	A1 B1, C1 A2, B2 C2 A3, B3 C3	B1, C1 A2, A1 B2 C2 A3, B3 C3	A1 B1, C1 A2, B2 C2 A3, B3 C3	A1 B1, C1 A2, B2 C2 A3, B3 C3
15624	A2 B2 B3 C3, A1 B1 C1 C2 A3	A1 B1 C1 C2, A2 B2, **A3 B3 C3**	**A1 B1 C1, A2 B2 C2, A3 B3 C3,**	**A1 B1 C1, A2 B2 C2, A3 B3 C3**
15603	A1 B2 C2, C1 B1 A2, **A3 B3 C3**	A1 B1 A2 C3, C1 B2 C2 A3 B3	C1 C3, A1 B2, B1 A2 C2, A3 B3	A1 B1 A2 B2 C2, C1 A3 B3 C3
15609	A1 B2, B1 C2, C1 C3, A3 B3, A2	A1 B1, C1 C3, B2 C2, A3 B3 A2	A1 A3, B1 C1, B2 C2, B3 C3 A2	A1 A3, B1 C1, B2 C2, B3 C3, A2

Note. This table provides an overview of the sorting tests. It shows the stimuli marked in groups by the participants. The participants numbers are seen in the left column. The upper panel is for Sequence 1, and the lower panel is for Sequence 2. The rows show the results for a given participant. The different sorting tests of the experiment are named in the headings of the columns. The stimuli marked in bold font show that the sorting is in accordance with the experimenter-defined class. The stimuli of the different marked groups are separated with commas.

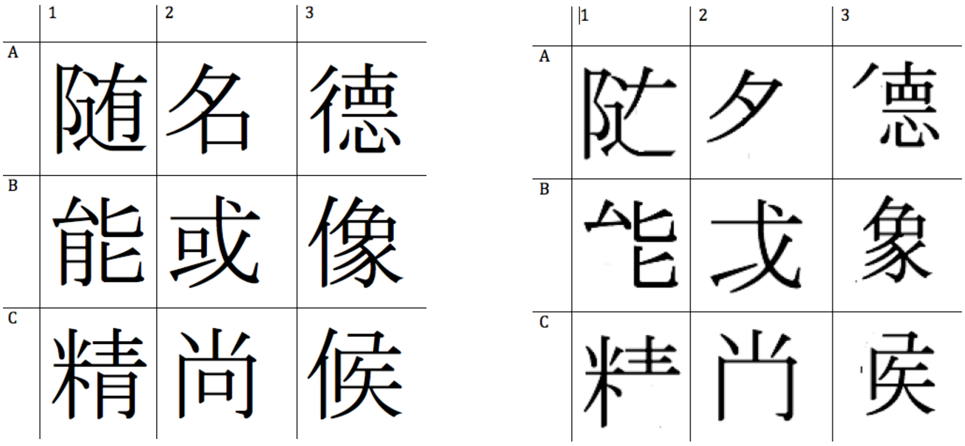
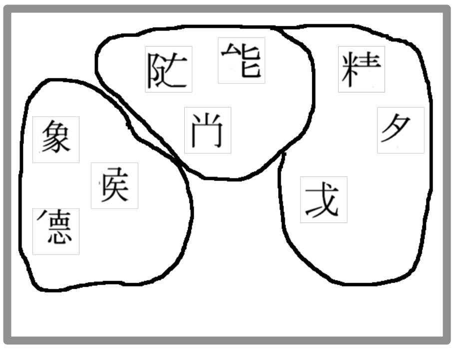
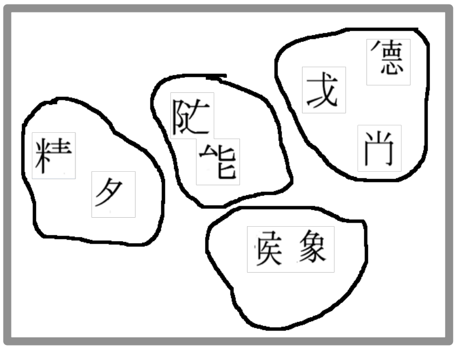


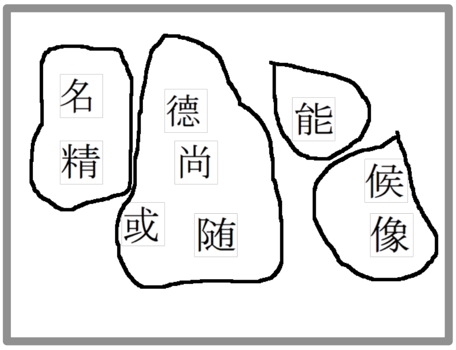
Figure 1. Stimuli used with 3-member classes. In the stimulus set in the left part of the figure, the characters have similar radicals for each class; this set is referred to as "with radicals" (WR). By manipulating the stimuli and removing all the similar radicals in each class, we created a new set of arbitrary stimuli, equal to the first set but without the radicals; this set is referred to as "without radicals" (WoR) and is depicted in the right part of the figure.



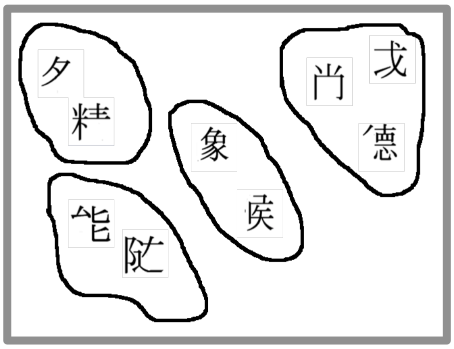
Participant 15623 Pre-SRT-WoR



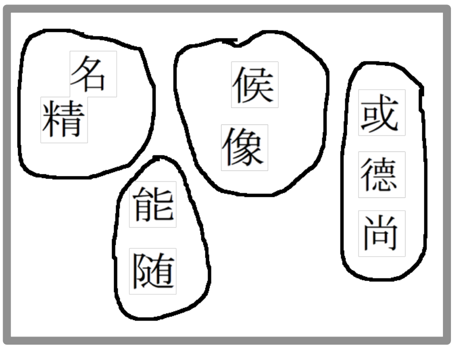
Participant 15615 Pre-SRT-WoR



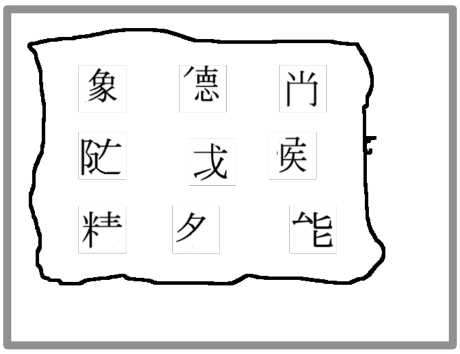
Participant 15615 Pre-SRT-WR



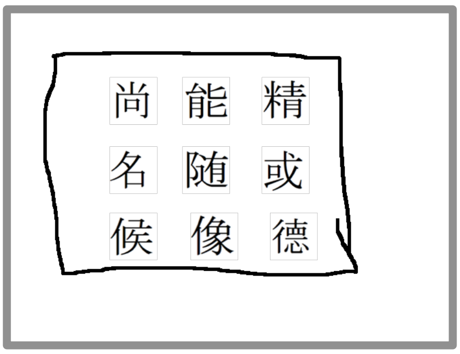
Participant 15615 Post-SRT-WoR



Participant 15615 Post-SRT-WR



Participant 15610 Pre-SRT-WoR



Participant 15610 Pre-SRT-WR

Figure 2. A reconstruction of the screen shots showing the performance of the different participants in the various sorting tests. The number of the participant and the sorting test type are written under each picture.

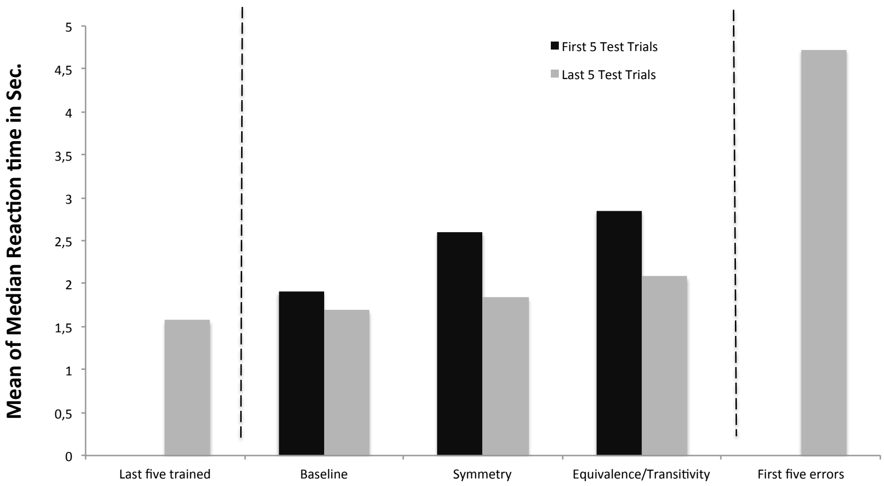


Figure 3. Group data of all twenty participants showing the mean median reaction times to comparison stimuli across the two sequences for the last five training trials (in the first section of the figure from the left), for the first five and last five test trials for baseline relations (Baseline), symmetry relations (Symmetry), and equivalence/transitivity relations (Equivalence/Transitivity) (in the middle section of the figure), and for all types of errors made in the test (in the third section of the figure).

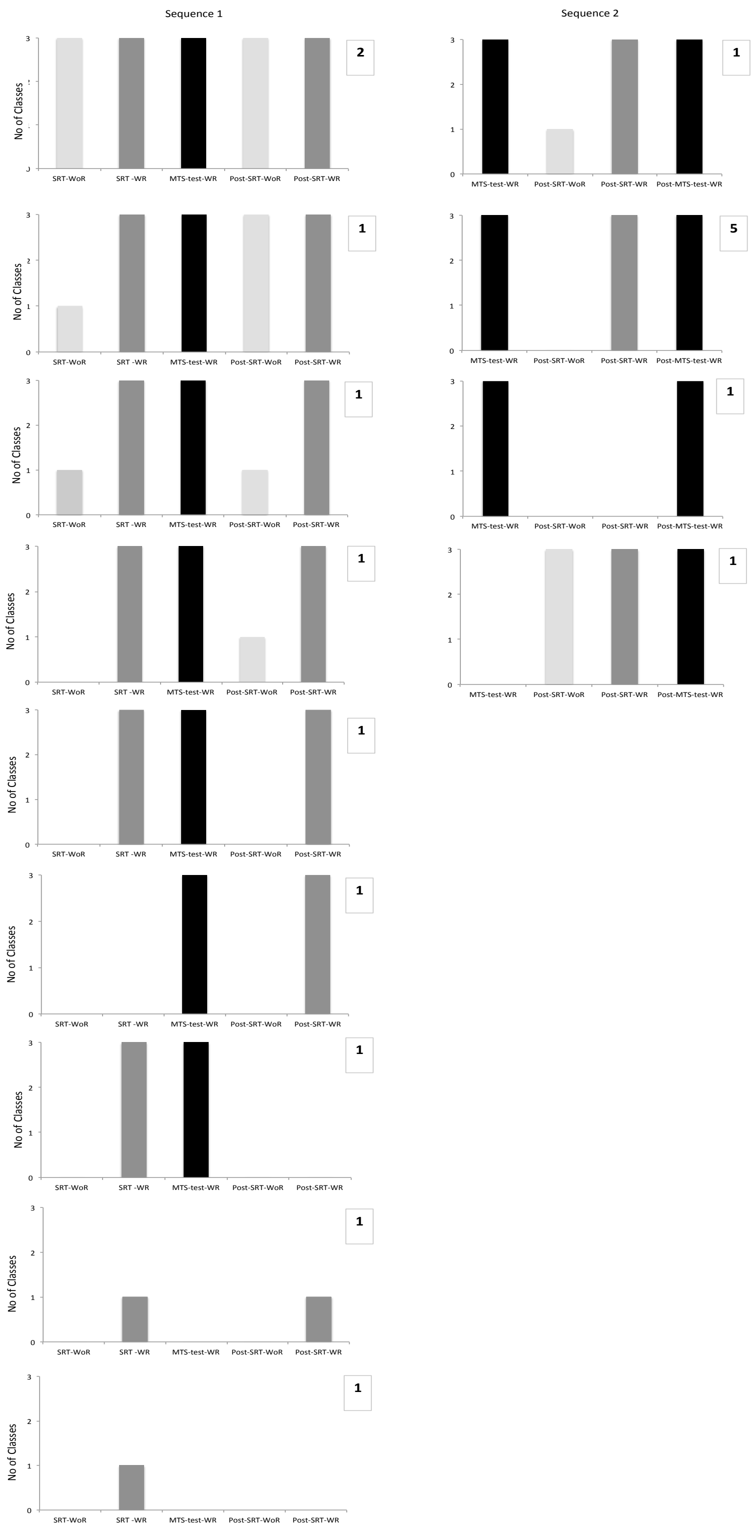


Figure 4. Number of classes that emerged in all test types and under each condition. The test types are written under the bars and follow the order of the tests in the sequences. The left column shows Sequence 1, and the right Sequence 2. The numbers in the boxes in the upper right corner of each panel report the number of participants showing each pattern across all test types. In Sequence 2, two participants did not form any experimenter-defined classes, so the number of participants showing the depicted patterns in Sequence 2 is eight.