Journal of the
Experimental Analysis of Behavior

SEAB

Journal of the Experimental Analysis of Behavior 2020, 1–17

On the effectiveness of including meaningful pictures in the formation of equivalence classes

Erik Arntzen and Justice Mensah

Oslo Metropolitan University, Norway

In three experiments, 165 adult participants were trained on 12 baseline conditional discriminations and tested for the formation of three 5-member equivalence classes (A→B→C→D→E). All experiments included two reference groups; the abstract (ABS) group was trained with all abstract stimuli and the picture (PIC) group with C-stimuli as meaningful pictorial stimuli but A, B, D, and E stimuli as abstract shapes. In Experiment 1, the color of the meaningful stimuli was manipulated. In the ABS, PIC, and black-and-white groups, 33.3%, 80%, and 93.3% formed equivalence classes, respectively. In Experiment 2, participants were exposed to a test block with and without trials that included C stimuli. For the groups with and without C trials in the test, 93.3% and 86.7% formed equivalence classes, respectively, compared to 20% in the ABS group. In Experiment 3, the number of meaningful pictures and their location in stimulus classes were manipulated. One group was trained with 3 pictures (C1/B2/D3, the 3-PIC) while the other groups had 2 pictures (C1/B2 and C1/D3, the 2-PIC). In the second test block for the ABS and PIC groups, 6.7% and 86% of the participants formed equivalence classes, respectively. For the 3-PIC and the 2-PIC groups, 66.7% and 50% of the participants formed equivalence classes, respectively. Results suggest that the effects of meaningful stimuli in equivalence classes (a) cannot be attributed to the use of colorful stimuli in previous studies, (b) occur during training and are not dependent on the presence of meaningful stimuli at test, and (c) are sensitive to stimulus location.

Key words: stimulus equivalence, delayed emergence, stimulus location, colorful stimuli, types of test trials, sorting performance, college students

---

Nearly 40 years ago, Sidman and Tailby (1982) described the defining features of equivalence class formation as reflexivity, symmetry, and transitivity (defined below). The most common arrangement in research on stimulus equivalence has been to train a certain number of conditional discriminations as a baseline for testing emergent relations. In an example with the minimal requirement of three stimulus sets (A, B, & C), the baseline conditional discriminations might consist of AB and BC relations. In this case, during training, A stimuli serve as sample stimuli in AB trials and C stimuli serve as comparisons in BC trials, while the B stimuli serve as both sample and comparison stimuli in AB and BC trials, respectively. In the subsequent test for emergent relations, participants are presented with BA, CB, AC, and CA trials. In accordance with terms adapted from mathematical set theory (Hrbacek & Jech, 1999, pp. 29–32), matching a B comparison to an A sample and a C comparison to a B sample is defined as symmetrical responding. Matching a C comparison to an A sample is defined as transitive responding, and matching an A comparison to a C sample is termed global equivalence. Reflexivity tests are rarely included in testing for emergent relations; however, such test trials would be AA, BB, and CC. Responding in accordance with stimulus equivalence is commonly observed in human participants. Even so, several variables have been shown to affect the likelihood of equivalence class formation (see Arntzen, 2012, for an overview).

One variable that influences the formation of equivalence classes is the type of stimuli used. For example, pronounceable nonsense

This research was funded by Oslo Metropolitan University.

The authors declare that there is no conflict of interest.

All procedures performed in studies involving human participants were in accordance with the ethical standards and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Data may be obtained upon request.

Address correspondence to: Erik Arntzen, Oslo Metropolitan University, Department of Behavioral Science, PO Box 4 St. Olavs Plass, 0130 Oslo, Norway. Email: erik.arntzen@equivalence.net

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

doi: 10.1002/jeab.579

© 2020 The Authors. *Journal of the Experimental Analysis of Behavior* published by Wiley Periodicals, Inc. on behalf of Society for the Experimental Analysis of Behavior

syllable (consonant-vowel-consonant [CVC] trigrams; Lyddy, Barnes-Holmes, & Hampson, 2000), familiar color-form compounds (Smeets & Barnes-Holmes, 2005), nameable stimuli (Bentall, Dickins, & Fox, 1993), pronounceable stimuli (Mandell & Sheen, 1994), rhyming stimuli (Randell & Remington, 2006), and meaningful pictures in both children and adults (Arntzen, 2004; Arntzen & Lian, 2010; Arntzen & Nikolaisen, 2011; Holth & Arntzen, 1998; O'Connor, Rafferty, Barnes-Holmes, & Barnes-Holmes, 2009) have been shown to substantially increase yields (i.e., the proportion of participants who meet criteria for equivalence class formation) on equivalence tests.

A recent line of research has explored the effects of including familiar color pictures as one of the stimuli in each potential class to model or mimic meaningfulness in a laboratory setting (Arntzen & Nartey, 2018; Arntzen, Nartey, & Fields, 2014, 2015a, 2018a, 2018b; Fields, Arntzen, Nartey, & Eilifsen, 2012; Mensah & Arntzen, 2017; Nartey, Arntzen, & Fields, 2014, 2015a, 2015b; Nedelcu, Fields, & Arntzen, 2015; Travis, Fields, & Arntzen, 2014; see also an overview in Fields & Arntzen, 2018). The common procedure in these experiments is to train adult participants on 12 conditional discriminations using a linear series (LS) training structure (A→B→C→D→E) and test for the formation of three 5-member equivalence classes. The experiments have addressed different research questions, but they have in all cases included two reference groups. In the abstract (ABS) group, all stimuli are abstract shapes. In the picture (PIC) group, the same abstract shapes serve as A, B, D, and E stimuli, but the abstract C stimuli, which occupy the middle position in the LS structure, are replaced with meaningful C stimuli: colorful pictures of items assumed to be familiar to human participants. Some of the studies have included groups that received training with the same stimuli as the ABS group, but also receive some form of preliminary training with the abstract C stimuli, including simple discrimination training (Fields et al., 2012), preliminary training with identity or arbitrary matching in combination with either simultaneous or delayed matching (Arntzen et al., 2014; Arntzen, Nartey et al., 2015), and preliminary training with meaningful pictures (Arntzen & Nartey, 2018). The main finding has been that inclusion of pictures as C stimuli substantially enhances equivalence class formation compared with the performance of the ABS group. Also, preliminary training with arbitrary C stimuli enhances equivalence class formation nearly as much as do pictures as C stimuli.

The goal of the present study was to address several questions that remain unanswered regarding the effects of meaningful stimuli on equivalence class formation. The first experiment addressed a difference between the abstract shapes and pictures used in previous studies: The pictorial stimuli in prior experiments have not only been meaningful but also colorful in comparison to the black and white abstract shapes. It has not been ruled out that the enhancing effect of the pictures could be a result of colorful stimuli being more salient than black and white shapes, instead of due to prior familiarity with the pictures. Thus, a comparison of colorful and black-and-white picture stimuli in the C position was undertaken to clarify the potential contribution of color to the enhancing effect of meaningful stimuli.

In the second experiment we asked about the role of the pictorial stimuli in the test for emergent relations. In the previous line of research, the enhancing effects of meaningful stimuli have been seen in the testing of emergent relations. However, the meaningful C stimuli have been included in both training and test trials and it is not clear if the effect depends on their presence at test or can be attributed to their presence in training alone. To separate the effects of meaningful stimuli in training from their effects during testing, we compared a PIC group that was exposed to meaningful C stimuli during both training and testing with a group that was exposed to meaningful stimuli during training but tested without C stimuli.

The third experiment examined unresolved issues regarding the location of meaningful pictures in stimulus classes. In most studies, conditions with meaningful stimuli have included one meaningful stimulus, C, in each putative class. However, previous findings have shown that equivalence class formation varies as a function of the location of the meaningful pictures in stimulus classes (Nartey et al., 2015b) and the number of stimulus classes containing meaningful stimuli (Mensah & Arntzen, 2017). In Nartey et al. (2015b), meaningful pictures were included in all stimulus classes but located in the A, B, C, D, or E position across different groups of participants. The main finding was that yields varied with location; specifically, pictures as C stimuli were more effective than any other location in five-member stimulus classes established using an LS training structure. Regarding the number of classes with meaningful C stimuli, Mensah and Arntzen (2017) compared the performance of four groups of participants; in addition to the usual ABS and PIC groups, one group had C-stimuli as pictures in two of three putative stimulus classes, and another group had a picture C-stimulus in only one of the three classes. The primary finding was that 80% and 50% of the participants in the PIC group and the group with two C stimuli as pictures responded in accordance with stimulus equivalence, respectively, compared to 13% for the group with one picture C stimulus and the ABS group. Together, these experiments suggest that the enhancing effects of meaningful stimuli depend on the location of the meaningful stimuli and the number of concurrently trained conditional discriminations that contain a meaningful stimulus. Specifically, when three stimulus classes were established concurrently, at least two of them had to contain a meaningful stimulus in order for enhancement to occur. To further explore these effects, in the present study we examined the effects of varying the location of the meaningful stimulus across stimulus classes (as for example as C1, B2, and D3) and the number of stimulus classes that contained a meaningful stimulus.

Two additional questions were asked in the present three experiments. First, in addition to measuring trials to mastery and yields on equivalence tests, we included performance on sorting tests as an additional measure to examine if the experimental manipulations affected this measure in a similar manner as equivalence test performance. The second question pertained to the possible increase in number of correct responses from the beginning to the end of testing (delayed emergence). Gradual emergence of equivalence classes, also described as delayed emergence because it has been observed that this change in response pattern can be abrupt, is a change in response patterns due to repeated testing (Sidman, 1994). Some of the papers in the line of the research mentioned above have reported delayed emergence. A typical finding is that a portion of participants shows immediate emergence (pass both test halves on the equivalence test) and another typically smaller portion responds in accordance with equivalence in the second half (Arntzen & Nartey, 2018; Arntzen et al., 2014; Arntzen, Nartey, et al., 2015; Arntzen et al., 2018a; Nartey et al., 2014, 2015a, 2015b) In sum, the nine papers showed delayed emergence of equivalence classes in 6.9%–23.1% of the participants. Also, one article reported delayed emergence as a function of nodes (Fields et al., 2012). However, the results were not presented as a function of trial types. Thus, in the present experiments we aimed to extend knowledge of delayed emergence by analyzing data on delayed emergence separately for different trial types (e.g., three-node transitivity vs. equivalence trials) in order to extend knowledge of delayed emergence by learning if it was connected with specific types of trials and varied as a function of number of nodes.

Experiment 1

As stated above, in the line of experiments which has explored the effects of pictures as C-stimuli on equivalence class formation, the pictures have been colorful while the abstract shapes have been black-and-white figures (e.g., Arntzen, Nartey, et al., 2015; Fields et al., 2012). To clarify the role of color versus meaningfulness in previous studies, Experiment 1 compared meaningful stimuli in color with meaningful stimuli in black-and-white while training 12 conditional discriminations and in the subsequent test for formation of three 5-member equivalence classes (A→B→C→D→E). Thus, we asked if equivalence class formation would vary as a function of the color of meaningful stimuli.

Method

Participants. Forty-five undergraduate students at the University of Ghana, Legon, served as participants in the experiment. Twenty-one males and 24 females aged from 15 to 28 years (M = 20.04, SD = 2.35) were recruited through personal contacts. None of the participants had previous knowledge about stimulus equivalence or conditional discrimination training. They were paid 50 Ghana Cedis (an equivalent of 11 US dollars) for participating in the experiment. The participants could withdraw from the experiment without any negative consequences. After finishing the experimental session, the participants were fully debriefed. The 45 participants were assigned randomly to three experimental groups: Abstract (the ABS group), C as meaningful pictures with color (the PIC group), and C as meaningful pictures in black-and-white (the PIC-as-BW group).

Apparatus and setting. Experimental sessions were conducted in a small lab room measuring approximately 7 m^2. A 15.6-in. screen HP ProBook 655 computer running Windows 8 equipped with a mouse and custom-made matching-to-sample (MTS) software was used to present stimuli and record participant responses.

Stimuli. The stimuli used in this experiment were abstract and meaningful pictures (see Fig. 1). The top section of the figure shows fifteen black and white abstract stimuli used in conditional discrimination training and testing for equivalence class formation in the ABS group. The middle section of the figure (PIC_1) shows three colorful meaningful picture-stimuli that replaced the C-abstract stimuli for the PIC group. The bottom section of the figure (PIC_2) shows a black-and-white version of the meaningful picture-stimuli in the middle section of the figure that replaced the C-abstract stimuli for the PIC-as-BW group. All the stimuli were presented on a white background.

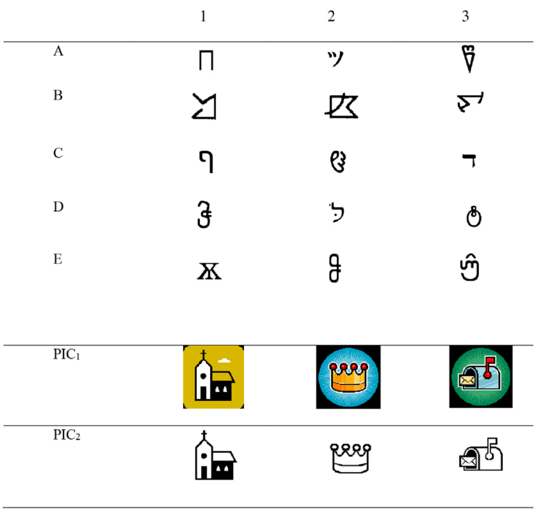

Fig. 1. The stimuli used in training conditional discrimination and testing for formation of equivalence classes. The top section shows the 15 abstract stimuli. The middle and bottom section show the colored and black-and-white meaningful pictures that replace the C-abstract stimuli in the respective picture groups. Stimuli A, B, C, D, and E and A, B, PIC_1 as C, D, and E are used in all three experiments as the reference groups (ABS and PIC). Stimuli A, B, PIC_2 as C, D, and E are used in Experiment 1. In Experiment 3, stimulus B2 was swapped with stimulus $PIC_1$2 and D3 was swapped with $PIC_1$3.

Procedure. Baseline training employed an LS training structure in which all baseline relations were trained concurrently. Specifically, AB, BC, CD, and DE baseline trials were presented mixed in one training block without replacement.

Instruction. At the beginning of the experiment, the following instruction in English was presented to the participant on the computer screen:

> In a moment a stimulus will appear in the middle of the screen. Click on this by using the computer mouse. Three stimuli will then appear in the three corners of the screen. Choose one of them by clicking on it with the mouse. If you choose the stimulus we have defined as correct, words like “very good”, “excellent” and so on will appear on the screen. If you press a wrong stimulus, the word “wrong” will appear on the screen. At the bottom of the screen, the number of correct responses you have made will be counted. During some stages of the experiment, the computer will NOT tell you if your choices are correct or wrong. However, based on what you have learned so far, you can get all of the tasks correct. Please do your best to get everything right. Thank you and good luck!

No further instructions were given before or after the experiment started.

Trial structure and contingencies. Each trial started with presentation of a sample stimulus in the center of the computer screen. Clicking on the stimulus with the mouse was followed by the presentation of three comparison stimuli displayed in three of the four corners of the computer screen. The location of the three comparison stimuli on the screen was assigned randomly. The selection of the correct comparison stimulus was followed by the presentation of words such as *correct, very good, super,* or *excellent.* The selection of an incorrect comparison stimulus was followed by the presentation of the word *wrong.* The duration of the programmed consequences was 1,000 ms followed by an intertrial interval of 500 ms. The mouse cursor was reset to the center of the screen prior to the presentation of a new sample stimulus.

Acquisition and maintenance of baseline relations. Table 1 presents an overview of baseline training. Baseline training was conducted in blocks of 60 trials including all the

Table 1

Sequence of training and testing

Experimental Phases	Trial Types	% Program Consequences	Number of Trials
Acquisition of baseline relations (All trial types presented randomly)			
1. Concurrent trials	A1B1, A2B2, A3B3, B1C1, B2C2, B3C3, C1D1, C2D2, C3D3, D1E1, D2E2, D3E3	100	60
2. Mixed trials (trials presented randomly)	A1B1, A2B2, A3B3, B1C1, B2C2, B3C3, C1D1, C2D2, C3D3, D1E1, D2E2, D3E3	75	60
3. Mixed trials (trials presented randomly)	A1B1, A2B2, A3B3, B1C1, B2C2, B3C3, C1D1, C2D2, C3D3, D1E1, D2E2, D3E3	50	60
4. Mixed trials (trials presented randomly)	A1B1, A2B2, A3B3, B1C1, B2C2, B3C3, C1D1, C2D2, C3D3, D1E1, D2E2, D3E3	0	60
Test 1 for emergent relations (trials presented randomly intermixed)			
Baseline trials	A1B1, A2B2, A3B3, B1C1, B2C2, B3C3, C1D1, C2D2, C3D3, D1E1, D2E2, D3E3	0	36
Symmetry trials	B1A1, B2A2, B3A3, C1B1, C2B2, C3B3, D1C1, D2C2, D3C3, E1D1, E2D2, E3D3	0	36
Transitivity	A1C1, A2C2, A3C3, A1D1, B1D1, A2D2, B2D2, A3D3, B3D3, A1E1, B1E1, C1E1, A2E2, B2E2, C2E2, A3E3, B3E3, C3E3	0	54
Equivalence	C1A1, D1A1, E1A1, C2A2, D2A2, E2A2, C3A3, D3A3, E3A3, D1B1, E1B1, D2B2, E2B2, D3B3, E3B3, E1C1, E2C2, E3C3	0	54
Test 2 for emergent relations (trials presented randomly intermixed)			
Baseline trials	A1B1, A2B2, A3B3, B1C1, B2C2, B3C3, C1D1, C2D2, C3D3, D1E1, D2E2, D3E3	0	36
Symmetry trials	B1A1, B2A2, B3A3, C1B1, C2B2, C3B3, D1C1, D2C2, D3C3, E1D1, E2D2, E3D3	0	36
Transitivity	A1C1, A2C2, A3C3, A1D1, B1D1, A2D2, B2D2, A3D3, B3D3, A1E1, B1E1, C1E1, A2E2, B2E2, C2E2, A3E3, B3E3, C3E3	0	54
Equivalence	C1A1, D1A1, E1A1, C2A2, D2A2, E2A2, C3A3, D3A3, E3A3, D1B1, E1B1, D2B2, E2B2, D3B3, E3B3, E1C1, E2C2, E3C3	0	54

Note. For the training phases, there are five trials of each relation per block, and for each set of the emergent relations test, there are three trials of each relation per block.

baseline relations (AB/BC/CD/DE). The trials presented in this block were A1/B1B2B3, A2/B1B2B3, A3/B1B2B3, B1/C1C2C3, B2/C1C2C3, B3/C1C2C3, C1/D1D2D3, C2/D1D2D3, C3/D1D2D3, D1/E1E2E3, D2/E1E2E3, and D3/E1E2E3, each appearing five times (the letter-number combination before the slash represents the sample, and the letter-number combinations after the slash represent the comparisons, with the correct comparison underlined). A mastery criterion of at least 90% correct comparison selection was required for the training of each relation. Participants therefore had to respond correctly in at least 54 out of the 60 trials required in the block. Performances below the 90% criterion led to a repetition of the block. Programmed consequences in this phase were delivered in 100% of all trials. Once the 90% criterion was achieved, participants proceeded to the maintenance phase, in which baseline training continued with reduced programmed consequences. The percentage of trials in a block that produced programmed consequences in the maintenance phase was reduced to 75%, 50%, and 0% in that order. A mastery criterion of at least 90% correct comparison selection in a trial block was required to progress to the next level of programmed consequence delivery. The trials that produced programmed consequences for each of the levels of programmed consequences were randomized in each block.

Emergent relations test blocks. The last block with no programmed consequences was followed by two test blocks for emergent relations that contained 180 trials each. Each test block consisted of 36 baseline, 36 symmetry, 54 transitivity, and 54 equivalence trials. The 180 trials in each test block were presented without programmed consequences. Equivalence class formation was defined as at least 90% experimenter-defined correct responding on all types of relations—baseline, symmetry, 1-node transitivity, 1-node equivalence, 2-node transitivity, 2-node equivalence, 3-node transitivity, and 3-node equivalence—in Test Blocks 1 and 2. Immediate emergence was defined as at least 90% correct responding in Test Block 1, while delayed emergence was defined as responding below 90% in Test Block 1 and at least 90% correct responding in Test Block 2. Furthermore, when looking at delayed emergence as a function of number of nodes, the criterion was that the responding at least 90% correct was achieved by the 1-node probes before the 2-node probes and 1-node probes and 2-node probes before 3-node probes.

Sorting test. Upon completion of the MTS-based test, participants were given plastic laminated cards with the same stimuli printed on the cards as they had been exposed to in MTS training and testing. Participants were instructed to put the cards into groups and their performances were recorded. The purpose of the sorting test was to provide an additional measure of class formation after MTS-based training and testing.

Results

Baseline acquisition. As shown in Figure 2, the PIC and the PIC-as-BW groups completed fewer training trials relative to the ABS group, $F(2, 42) = 3.96$, $p = .03$. Participants in these groups required a mean of 600, 600, and 728 trials, respectively, before the test for emergent relations.

Equivalence class formation. Figure 3 shows the percentage of participants who formed equivalence classes in the immediate (Test Block 1) and delayed tests (Test Block 2) for emergent relations and sorted the cards according to the experimenter-defined classes. In the ABS group, 13.33% and 33.33% of the

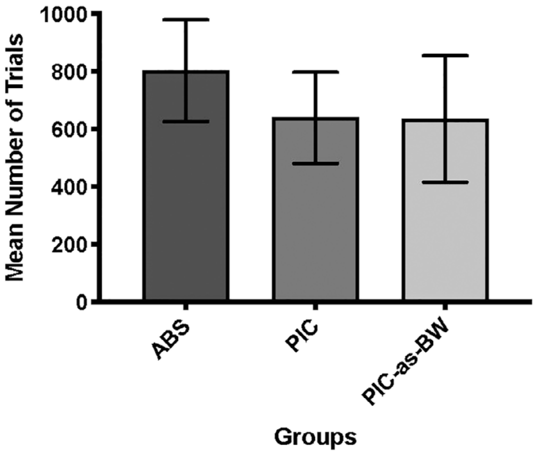

Fig. 2. Mean number of trials needed to acquire baseline relations for participants in each of the experimenter groups in Experiment 1. Error bars show standard error of the mean. ABS = All stimuli are abstract, PIC = C stimuli are colored pictures, and PIC-as-BW = C stimuli are black-and-white pictures.

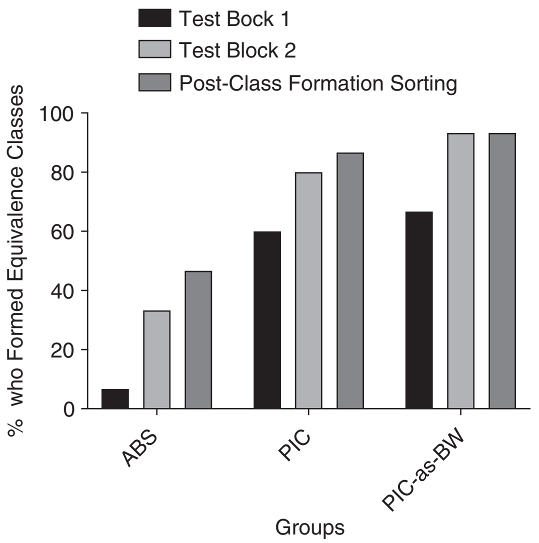

Fig. 3. The percentage of participants who showed class formation in immediate and delayed emergence relations tests, and postclass formation sorting test in Experiment 1. Each cluster of bars represents yields for a specific group, with the leftmost bar showing data for the first test block, the middle bar showing data for the second test block, and rightmost bar showing yields on the postclass formation sorting test. ABS = All stimuli are abstract, PIC = C stimuli are colored pictures, and BW-as-PIC = C stimuli are black-and-white pictures.

participants formed classes in Test Blocks 1 and 2, respectively. In the PIC group, 66.7% and 80% of the participants formed classes in Test Blocks 1 and 2, respectively, and in the PIC-as-BW group, 66.67% and 93.33% of the participants formed classes in Test Blocks 1 and 2, respectively. Fisher's Exact Tests showed that in Test Block 2, class formation yields were significantly higher for PIC relative to ABS, p = .025, 95% CI [0.03, 0.61], and PIC-as-BW relative to ABS, p = .002, 95% CI [0.00, 0.3], but not for PIC relative to PIC-as-BW, p = .60, 95% CI [0.02, 2.22]. On the sorting task, 46.67%, 86.67%, and 93.33% of the participants in the ABS, the PIC, and the PIC-as-BW groups, respectively, sorted the cards according to the experimenter-defined classes. Fisher's Exact Tests showed that performance consistent with experimenter-defined classes was significantly higher for PIC compared to ABS, p = .021, 95% CI [0.02, 0.56] and performance for PIC-as-BW relative to ABS, p = .005, 95% CI [0.0, 0.4]. Furthermore, a Fisher's Exact Test showed no significant difference in experimenter-defined class formation, and PIC compared to PIC-as-BW, p = 1.0, 95% CI [0.03, 4.47].

Delayed emergence. Figure 4 shows a more detailed analysis of the performance of participants who did not respond in accordance with equivalence in Test Block 1 but formed equivalence classes in Test Block 2. Twenty of the total 45 participants responded with equivalence in both test blocks. In addition, nine participants (three participants in the ABS group, two participants in the PIC group, and four participants the BW-as-PIC group did not show immediate equivalence class formation in Test Block 1, but responded in accordance with stimulus equivalence in Test Block 2 (delayed emergence). As shown in Figure 4, the number of participants showing delayed emergence was lowest (two of nine) for symmetry relations and highest (eight of nine) for 3-node equivalence relations. Figure 4 shows that three of the participants show a response pattern that is in accordance with criterion of showing delayed emergence as function of number of nodes.

Discussion

We asked if the combination of color and content of the stimuli used as meaningful pictures in previous studies could have affected the results due to the salience of color stimuli. However, both conditions (colorful pictures and black-and-white pictures), replicated the earlier enhancing effect of familiar pictures on equivalence class formation. This finding suggests that enhanced discriminability is not dependent on the color but rather on the content of the stimuli relative to abstract shapes. For practical purposes, when arranging and conducting a procedure which is going to be effective in producing equivalence class formation, it is helpful for stimuli in one stimulus set to be meaningful but not necessarily in color.

Experiment 2

Studies on the relatedness of stimuli in equivalence classes have used meaningful pictures during MTS training, but not included meaningful pictures in MTS emergent relations testing (Bortoloti & de Rose, 2009, 2012; Bortoloti, Rodrigues, Cortez, Pimentel, & de Rose, 2013; de Almeida & de Rose, 2015). These studies have generally used a training

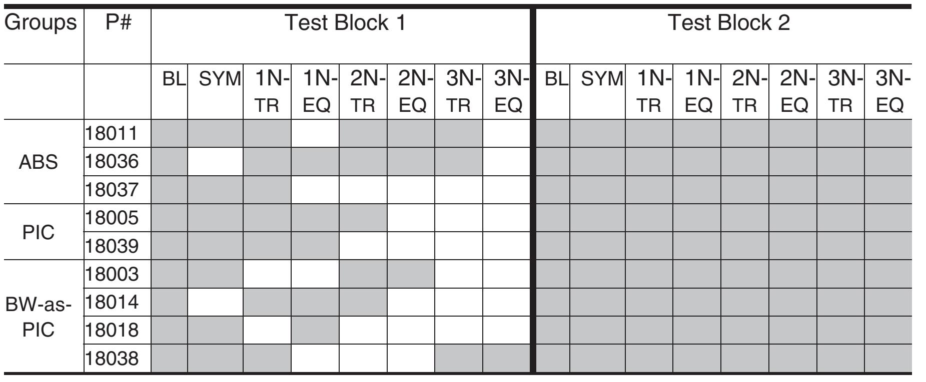

Fig. 4. The figure shows all participants across different groups in Experiment 1 who did not respond in accordance with stimulus equivalence in Test Block 1. The filled squares show the relations in which participants responded in accordance with experimenter-defined criterion, while the open squares show the relations in which participants did not respond in accordance with experimenter-defined criterion. ABS = All stimuli are abstract, PIC = C stimuli are colored pictures, and BW-as-PIC = C stimuli are black-and-white pictures.

structure in which the trained relations are BA, AC, and CD (B←A→C→D). The B, C, and D stimuli are abstract shapes, while the A stimuli are pictures of faces showing emotions (happiness, anger, and neutrality). The test includes BD and DB trials, so the pictures of faces (A stimuli) are not presented in test trials. By contrast, studies on equivalence class formation as a function of meaningful pictures have used meaningful pictures during both MTS training and emergent relations testing (e.g., Arntzen, Nartey et al., 2015). In these experiments, all trial types have been included in the test blocks for emergent relations and obtained yields have been 0–20% for the ABS groups and 70–80% for the PIC groups (e.g., Arntzen, Nartey et al., 2015; Mensah & Arntzen, 2017). It is possible that the high yields are due to the inclusion of meaningful pictures in some test trials (i.e., all trials with C stimuli). Thus, Experiment 2 explored the effect of including or excluding trials with meaningful C stimuli in testing for the formation of three 5-member equivalence classes (A→B→C→D→E).

Method

Participants. The experiment included 15 male and 30 female undergraduate students aged from 17 to 29 years (M = 22.02; SD = 2.70) who were recruited in the same manner and given the same information as in Experiment 1.

The participants were assigned randomly to three experimental groups: ABS and PIC as in Experiment 1, and the Wo-PIC group, which received the same training as the PIC group but received the test for emergent relations without C trials.

Procedure. Apparatus, setting, relations trained, and training procedures were the same as in Experiment 1. The ABS group received training with the A, B, C, D, and E sets of stimuli shown in Figure 1, and C stimuli were replaced with the PIC_1 stimuli for the PIC and Wo-PIC groups; the PIC_2 stimuli were not used in this experiment. The testing procedure differed from Experiment 1 only in that the Wo-PIC group had no test trials with the C stimuli (i.e., no AC, BC, CD, or EC) in the emergent relations test. Therefore, for this group, each of the emergent relations test blocks consisted of 18 baseline, 18 symmetry, 36 transitivity, and 36 equivalence trials totaling 108 test trials.

Results

Baseline acquisition. As shown in Figure 5, the mean number of trials to baseline mastery was 900, 660, and 780 for the ABS, the PIC, and the Wo-PIC groups, respectively; the difference between groups was not significant, F (2,42) = 1.98, p = .15.

Equivalence class formation. Figure 6 shows the percentage of participants who formed equivalence classes in the first and second test

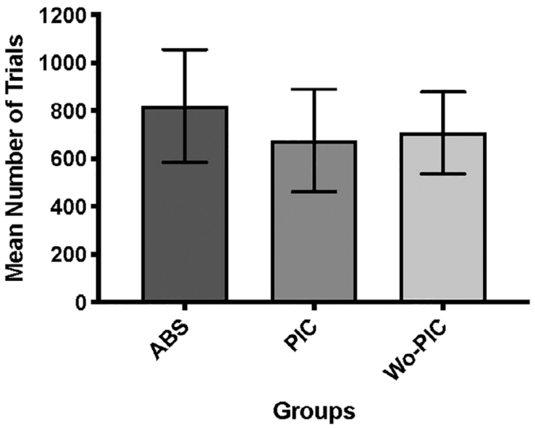

Fig. 5. Mean number of trials needed to acquire baseline relations for participants in each of the experimenter groups in Experiment 2. Error bars show standard error of the mean. ABS = All stimuli are abstract, PIC = C stimuli are pictures and included in both training and testing, Wo-PIC = C trials are excluded during testing.

blocks and sorted the cards according to the experimenter-defined classes. In the ABS group, 6.67% and 20% of the participants showed equivalence class formation in Test Blocks 1 and 2, respectively, as did 60% and 93.3% of the participants in the PIC group, respectively, and 60% and 93.3% of participants in the Wo-PIC group. Fisher's Exact Tests showed that in Test Block 2, class formation yields were significantly higher for both PIC relative to ABS and Wo-PIC relative to ABS, p = .00, 95% CI [0.0, 0.18) but not for PIC relative to Wo-PIC, p = 1, 95% [0.04, 20,23]. For the sorting task, 46.67%, 93.33%, and 93.33% of the participants in the ABS, the PIC, and the Wo-PIC groups, respectively, sorted the cards according to the experimenter-defined classes. Fisher's Exact Tests showed that experimenter-defined class formation performance was significantly higher for both PIC compared to ABS and Wo-PIC compared to ABS, p = .014, 95% CI [0.0, 0.53]. Furthermore, a Fisher's Exact Test showed no significant difference in experimenter-defined class formation performance for PIC relative to Wo-PIC, p = 1.0, 95% CI [0.04, 20.23].

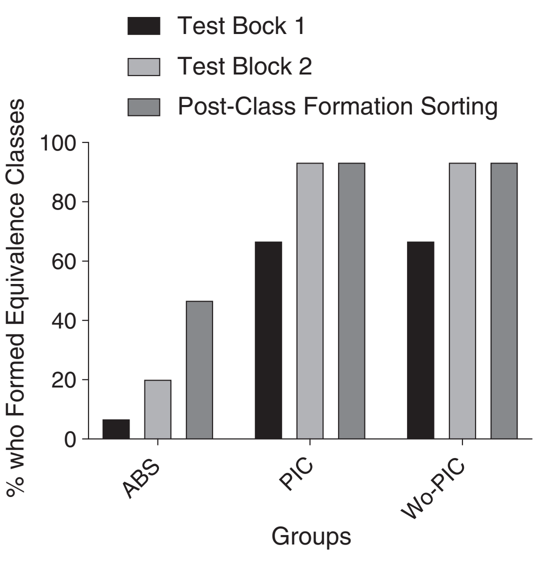

Fig. 6. The percentage of participants who showed class formation in immediate and delayed emergence relations tests, and postclass formation sorting test in Experiment 2. Each cluster of bars represents yields for a specific group, with the leftmost bar showing data for the first test block, the middle bar showing data for the second test block, and rightmost bar showing yields on the postclass formation sorting test. ABS = All stimuli are abstract, PIC = C stimuli are pictures and included in both training and testing, Wo-PIC = C related relations are excluded during testing.

Delayed emergence. As shown in Figure 6, 42.22% of the participants showed equivalence class formation in both test blocks. None of the participants responded in accordance with equivalence in Test Block 1 but not in second test block. Eleven participants (two participants in the ABS group, five participants in the PIC group, and three participants in the Wo-PIC group who did not form equivalence classes in the immediate test for emergent relations, formed classes in the second test block (see Fig. 7). Three out of four participants in the Wo-PIC group did not respond according the experimenter-defined criterion in baseline trials. The number of participants showing delayed emergence was lowest (2 of 10) for symmetry relations and highest (9 of 10) for 3-node equivalence relations. One participant, P17005, only responded correctly on baseline trials in the first test block. P17019 and P17033 showed delayed emergence of equivalence classes as a function of number of nodes.

Discussion

Experiment 2 extended previous findings by examining the effects of eliminating meaningful pictures from the MTS emergent relations test after including them in the MTS training.

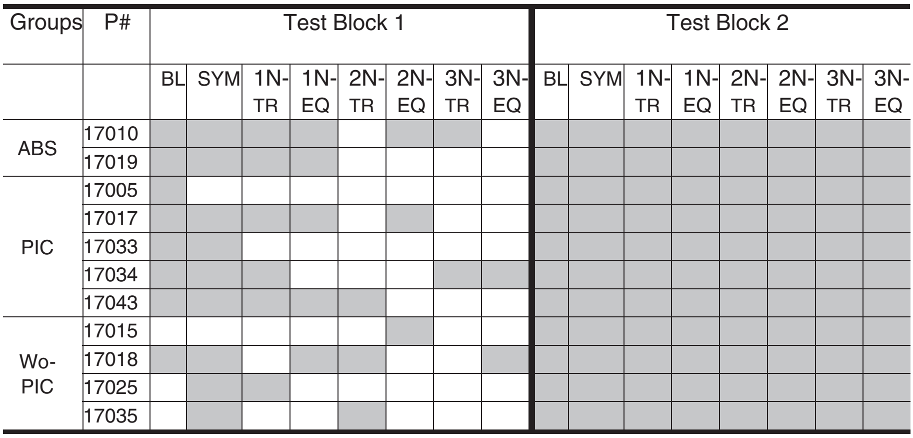

Fig. 7. The figure shows all participants in Experiment 2 who did not respond in accordance with stimulus equivalence in Test Block 1. The filled squares show the relations in which participants responded in accordance with experimenter-defined criterion, while the open squares show the relations in which participants did not respond in accordance with experimenter-defined criterion. ABS = All stimuli are abstract, PIC = C stimuli are pictures and included in both training and testing, Wo-PIC = C trials are excluded during testing.

Results showed no significant difference between class formation yields for participants with or without meaningful pictures included in the MTS emergent relations test. Thus, we argue that it is the training of baseline relations with meaningful pictures that enhances discriminability during the establishment of necessary conditional discriminations and not the mere presence of meaningful pictures in the testing of emergent relations. As an alternative interpretation to the discriminability of meaningful pictures, it is possible that meaningful stimuli enhance mediation behavior. Mediating behavior has been suggested as important to fill the gap between sample offset and comparison onset in delayed matching-to-sample procedures (e.g., Arntzen, 2006; Arntzen et al., 2014). Other studies have shown how verbal behavior may influence the formation of equivalence classes (e.g., Jennings & Miguel, 2017). Further experiments including talk-aloud procedures might be helpful for clarifying the role of mediating behavior induced by meaningful pictures (Vie & Arntzen, 2017).

Experimental sessions within stimulus equivalence research can be quite lengthy. The results from Experiment 2 show that when training a high number of conditional discriminations (in this case, 12 conditional discriminations including meaningful pictures as C stimuli), the number of trials in a test block can be reduced by excluding all trials with C stimuli, thereby reducing the duration of the experimental sessions.

Experiment 3

Previous research suggests that meaningful pictures presented as the C set in a five-member class have a greater effect on equivalence yields compared to pictures as the A-, B-, D-, or E-set (Nartey et al., 2015b), and that presenting a pictorial C stimulus in all concurrently taught stimulus classes is more effective than presenting fewer (Mensah & Arntzen, 2017). To extend these findings, in Experiment 3 we studied the effects of including a picture in each class but in different sets. As before, we trained 12 conditional discriminations and tested for three 5-member equivalence classes (A→B→C→D→E), and we included groups in which the location of the meaningful stimulus varied across classes and either two or three classes contained a meaningful stimulus.

Method

Participants. The participants were 75 undergraduate students (46 males and 29 females) recruited through personal contacts from the University of Ghana, Legon. Their ages ranged from 16 to 29 years (*M* = 20.12, *SD* = 2.37).

The 75 participants were assigned randomly to five experimental groups: Abstract (the ABS group), **C1, C2,** and **C3** stimuli as pictures (the PIC group), **C1**, **B2**, and **D3** stimuli as pictures (the 3-PIC), and two groups with pictures as **C1** and **B2** or **D3** stimuli (the 2-PIC groups).

Procedure. Apparatus, setting, stimuli, relations trained, and training procedures were the same as in Experiment 1. As before, the ABS group received training with the A, B, C, D, and E sets shown in Figure 1, and the PIC group received training with C_1 instead of C stimuli. For the 3-PIC group, $PIC_1 1$, $PIC_1 2$, and $PIC_1 3$ stimuli replaced C1, B2, and D3, and in the two 2-PIC groups, $PIC_1 1$ and $PIC_1 2$ stimuli replaced C1 and B2, and C1 and D3, respectively. The PIC_2 stimuli were not used in this experiment.

Results

Baseline acquisition. Figure 8 shows the mean number of trials needed to acquire baseline relations in each experimental group. Participants required a mean number of 700, 588, 669, and 574 trials to acquire baseline relations in the ABS, the PIC, the 3-PIC, and the 2-PIC groups (the data for the two 2-PIC groups C1 and B2 or D3 as pictures have been averaged because they did not differ and both included one picture in addition to C1). A one-way ANOVA showed no significant difference between the experimental groups, $F(3, 56) = 1.62$, $p = .33$.

Equivalence class formation. Figure 9 shows the percentage of participants who responded in accordance with equivalence in Test Block

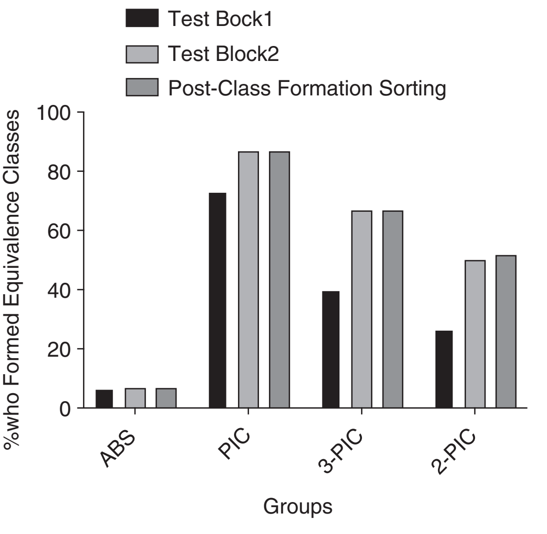

Fig. 9. The percentage of participants who showed class formation in Test Block 1, Test Block 2, and postclass formation sorting test in Experiment 3. Each cluster of bars represents yields for a specific group, with the leftmost bar showing data for the first test block, the middle bar showing data for the second test block, and rightmost bar showing yields on the postclass formation sorting test. ABS = All stimuli are abstract, PIC = C stimuli are pictures, 3-PIC = C1, B2, and D3 stimuli are pictures, 2-PIC = C1 and B2 or D3 stimuli are pictures.

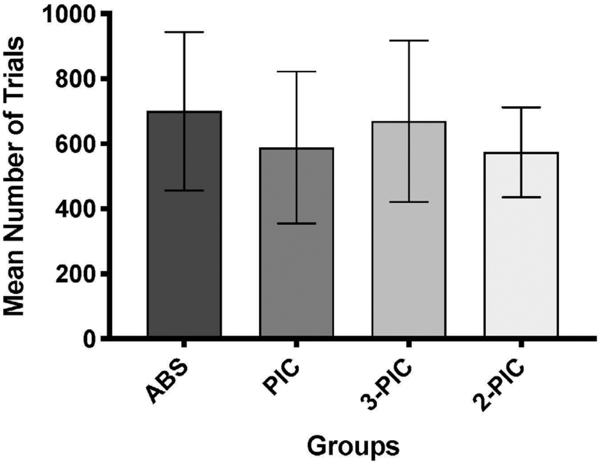

Fig. 8. Mean number of trials needed to acquire baseline relations for participants in each of the experimenter groups in Experiment 3. Error bars show standard error of the mean. ABS = All stimuli are abstract, PIC = All three C stimuli are pictures, 3-PIC = C1, B2, and D3 stimuli are pictures, 2-PIC = C1 and B2 or D3 stimuli are pictures.

1 (immediate emergence) and Test Block 2 (delayed emergence) and sorted the cards according to the experimenter-defined classes. For the ABS group, 6.7% of the participants responded in accordance with equivalence in both the first and second test blocks. In the PIC group, 73.3% and 86.7% of the participants responded in accordance with equivalence in the first and second test blocks, respectively. For the remaining groups, these respective yields were 40% and 66.7% in the 3-PIC group, 26.7% and 50% in the two 2-PIC groups. In Test Block 2, the Fisher's Exact Tests showed that class formation yields were significantly higher for PIC compared to ABS, p = .0001, 95% CI [6.66, 938.6], 3-PIC compared to ABS, p = .0017, 95% [2.93, 298.8], and 2-PIC compared to ABS, p = .0071, 95% [1.68, 148.5]. The results also show that class formation yields were significantly higher for PIC compared to C1/B2-as-PIC and C1/D3, p = .02, 95% CI [1.34, 31.70]. The Fisher's Exact Tests also showed no significant difference in class formation yields for PIC and 3-PIC, p = .39, 95% CI [0.46, 18.27] and 3-PIC and 2-PIC, p = .35, 95% CI [0.59, 6.89]. For the sorting task, the number of participants who sorted according to experimenter-defined classes corresponds with the performance in Test Block 2.

Delayed emergence. As shown in Figure 10, 36% of the participants responded in accordance with stimulus equivalence in both test blocks. Eleven participants, two participants in the PIC group, four participants in the 3-PIC group, three participants in the 2-PIC group with C1 and B2 as pictures, and two participants in the 2-PIC group with C1 and D3 as pictures who did not respond in accordance with stimulus equivalence in Test Block 1 did so in Test Block 2. One participant, P19004, did not respond correctly in baseline trials and one participant, P19012, only responded correctly on baseline trials in Test Block 1. The number of participants showing delayed emergence was lowest (2 of 10) for symmetry relations and highest (10 of 11) for 3-node transitivity relations. P19031, P19071, and P19081 showed delayed emergence of either both types of 3-node relations or one 3-node relation. The other five participants (P19004, P19007, P19016, P19012, P19025, P19048, and P19049) showed delayed emergence of both low and high nodal number. Two participants, P19031 and P19081, showed delayed emergence of equivalence classes as a function of number of nodes.

Discussion

In the previous line of research using five sets of stimuli, A, B, C, D, and E, the meaningful pictorial stimuli have always been presented in the same set (e.g., Arntzen & Nartey, 2018), specifically as set-C samples and comparisons (C1, C2, C3). According to Fields

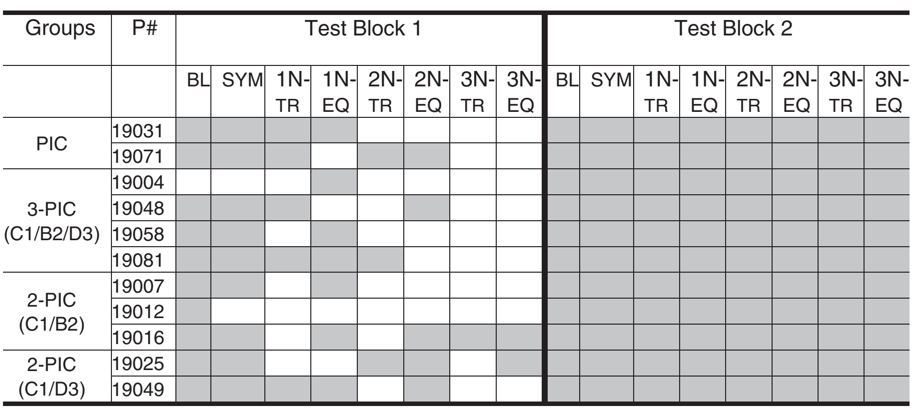

Fig. 10. All participants in Experiment 3 who did not respond in accordance with stimulus equivalence in Test Block 1. The filled squares show the relations in which participants responded in accordance with experimenter-defined criterion, while the open squares show the relations in which participants did not respond in accordance with experimenter-defined criterion. ABS = All stimuli are abstract, PIC = C stimuli are pictures, 3-PIC = C1, B2, and D3 stimuli are pictures, and 2-PIC = C1 and B2 or D3 stimuli are pictures.

et al. (2012) and Travis et al. (2014), the class-enhancing effect of the meaningful stimuli can be attributed to their being members of an already existing equivalence class before the experiment. Therefore, the meaningful stimuli can function as discriminative stimuli for class formation in a set of abstract stimuli. Fields et al. (2012) suggested that the formation of an equivalence class consisting of four abstract stimuli and a meaningful stimulus indicates an expansion of a preestablished class of which the meaningful stimulus is already part, rather than a "new" equivalence class of five members.

The results showed that when the meaningful pictures were all C-set stimuli (samples and comparisons), the yields in the first test block were more than 30% higher than when the three meaningful pictures were part of different sets (C1/B2/D3), but the difference was smaller in the second test block (see the discussion on delayed emergence below). When all three meaningful pictures are presented as C stimuli in a five-member class using an LS training structure, they are trained to three B and three D stimuli. Both B and D stimuli are nodes that serve both as samples and comparisons in different trial types. When the three meaningful pictures were presented in different sets to the 3-PIC group, two of them were also trained to singles (a stimulus trained to one other stimulus only); that is, B2 was trained to A2, which served only as a sample, and D3 was trained to E3, which only served as a comparison. Further research with a larger number of stimulus sets is needed to determine if the smaller yields in this group are related to the meaningful stimuli being members of different stimulus sets per se, or if they are related to some of the meaningful stimuli being trained to singles.

The low yields of equivalence class formation for the two 2-PIC groups compared to the group with pictorial stimuli in three classes can be attributed to the absence of a meaningful stimulus in one of their experimental classes.

General Discussion

Across all three experiments, the inclusion of a meaningful stimulus in a group of abstract stimuli produced equivalence yields that were significantly higher than when the group of stimuli contained only abstract stimuli. Also, for participants who did not form equivalence classes, the proportion of probe trials that were consistent with experimenter-defined classes was higher for all picture groups compared to the abstract group in all three experiments. These findings replicate those previously reported by Arntzen et al. (2014); Arntzen, Nartey et al. (2015); Fields et al. (2012); Mensah & Arntzen (2017); and Nartey et al. (2015a, 2015b). We did not find any differences in yields when comparing colorful pictorial stimuli and black-and-white pictorial stimuli (Experiment 1) or the inclusion and exclusion of test trials including meaningful pictures as C stimuli (Experiment 2). Furthermore, Experiment 3 showed that when the meaningful pictures were located in different stimulus sets (not as C stimuli only), the yields were little reduced compared to when the meaningful pictures served as stimuli in the C set. However, yields were reduced substantially when only two comparisons in different sets were meaningful pictures, replicating previous findings (Mensah & Arntzen, 2017).

Meaningful Stimuli

Fields and Arntzen (2018) mentioned different properties that are characteristic for meaningful stimuli—hedonic valence, denotative (definitional) and connotative (evaluative) features. As pointed out in the introduction, different experiments have shown how meaningful stimuli that acquired stimulus control functions (discriminative and conditional stimuli) can influence equivalence class formation. This enhancing effect is also, as discussed in Experiment 3, found when meaningful stimuli serve as members of other stimulus classes. Taken together, the present three experiments clarified some unresolved issues regarding the effect of including meaningful stimuli in stimulus sets with abstract shapes. Fields and Arntzen called for experiments to clarify if the enhancement of class formation is related to a combination of hedonic, denotative and connotative properties and the acquired stimulus control functions.

Sorting

Data on the sorting tests for the ABS and the PIC groups in the three experiments in the present study replicate findings in previous studies with LS (Arntzen, Granmo, & Fields, 2017; Arntzen, Norbom, & Fields, 2015b; Fields, Arntzen, & Moksness, 2014) and MTO (Rustad Bevolden & Arntzen, 2018) training structures. Specifically, all participants who formed equivalence classes in the MTS-based emergent relation test sorted the cards according with experimenter-defined classes. Furthermore, some participants who failed to form classes in the MTS-based emergent relations test sorted the cards according to experimenter-defined classes.

When we compare the sorting results from the three experiments in the present study for different experimental groups other than the reference groups (the ABS and the PIC groups) with previous research mentioned above, the number of participants who sorted correctly was higher for the PIC-as-BW group (Experiment 1) and was the same for the Wo-PIC group (Experiment 2) as the PIC groups. Furthermore, correct sorting for the C1/B2/D3-as-PIC, the C1/B2-as-PIC and the C1/D3-as-PIC groups decreased in a graded manner compared to the performance for the PIC group (Experiment 3). These results suggest that the number of pictures as C stimuli in the classes and the location of pictures in different stimulus sets influence sorting performance while the color of the pictures and tests without C trials do not. In other words, there is a high correlation between MTS performance and performance on sorting tests and both are influenced by the same variables.

There are several practical aspects to the use of sorting tests. First, it takes much less time to do the sorting test than a full MTS test (Arntzen et al., 2017), and second, sorting tests are easy to administer (Arntzen, Norbom et al., 2015). Third, sorting tests are useful in settings with large numbers of people (Varelas & Fields, 2017). Finally, sorting tests can be used with small children (Barron, Leslie, & Smyth, 2018).

Delayed Emergence of Equivalence Classes

Equivalence class formation is demonstrated when all test trials for emergent relations are in accordance with the experimenter-defined classes. However, many reports within the literature on stimulus equivalence have shown an increase in responding according to experimenter-defined classes as a function of repeated test trials or continued testing (e.g., Arntzen & Holth, 2000; Doughty, Leake, & Stoudemire, 2014; Fields, Adams, Verhave, & Newman, 1990; Holth & Arntzen, 1998; Imam & Blanche, 2013; Kato, de Rose, & Faleiros, 2008; Sidman, Kirk, & Willson-Morris, 1985).

Some researchers have claimed that delayed emergence does not occur often (Dube & McIlvane, 1996) while others suggest it is a relatively common finding (Pilgrim, 2015). In the present study, delayed emergence was observed for some participants in all groups, including the ABS group (Experiments 1 & 2). The analysis showed that a total of 31 out of 165 participants across three experiments showed delayed emergence and seven of these participants showed a pattern of delayed emergence as a function of number of nodes. This finding is consistent with previous findings showing that number of nodes could influence delayed emergence of equivalence classes (e.g., Fields et al., 1990; Kennedy, 1991; Moss-Lourenco & Fields, 2011). Twenty-three out of the 30 participants showed a mixed pattern of delayed emergence and not as function of number of nodes. Furthermore, the analysis of separate trial types extends previous findings and shows that it could be differences in performance on transitivity versus equivalence on 1-node, 2-node, and 3-node trials.

In the present experiments, baseline trials without programmed consequences were interspersed in the test. It is possible that the presentation of baseline trials in a test block plays a role in delayed emergence of equivalence classes. For example, Sidman (1994) emphasized how the consistent sample–comparison relations are established in the conditional discrimination and that this consistency from many other relations should enable performance according with experimenter-defined classes during testing, and therefore lead to delayed emergence. Based on this argument, delayed emergence should be studied as function of testing for equivalence with and without baseline trials in test blocks. The prediction is that without the consistent sample–comparison relations in the test, the number of participants showing delayed emergence should be low.

A related mechanism for delayed emergence is put forward by Fields and colleagues

(Fields & Moss, 2007; Moss-Lourenco & Fields, 2011). They argue that delayed emergence of equivalence classes may occur because the between-class discriminations during training are not fully established, even if the mastery criterion before testing is met. Experimentally, this suggestion could be tested by including an extended phase with direct training of between-set discrimination trials and examine if cases of delayed emergence are reduced.

Summary and Implications

Taken together, all three experiments replicate and extend previous research on the enhancing effect on equivalence class formation of including meaningful pictures in a stimulus set with abstract shapes. The results of the experiments clarify important variables that are responsible for this enhancing effect. Some of main implications of the three experiments are: (1) when training and testing for three potential classes, the enhancing effect of equivalence class formation is optimized with a meaningful picture in all three classes as nodes but not necessarily located as the middle node in each class. It is not clear what the effect would be with pictures located in some classes as nodes and as singles in other classes. (2) Experiments arranged to show the emergence of large equivalence classes can be quite time consuming. The results showing the high correspondence of performance on sorting tests and MTS test, and the finding that trials including meaningful stimuli in test trials are not necessary to enhance the formation of equivalence classes as long as meaningful stimuli are included in baseline conditional discriminations, could be significant in applied settings. Both the use of sorting tests as a measurement for partitioning of classes and reducing the number of test trials in the MTS tests will reduce time spent in an experiment. (3) The mechanism of delayed emergence of equivalence could be studied by manipulating training and testing arrangements.

References

Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *The Psychological Record, 54,* 275–291. doi:https://doi.org/10.1007/BF03395474

Arntzen, E. (2006). Delayed matching to sample and stimulus equivalence: Probability of responding in accord with equivalence as a function of different delays. *The Psychological Record, 56,* 135–167. doi:https://doi.org/10.1007/BF03395541

Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: Methodological issues. *European Journal of Behavior Analysis, 13,* 123–135. doi:https://doi.org/10.1080/15021149.2012.11434412

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, 67,* 81–96. doi:https://doi.org/10.1007/s40732-016-0209-9

Arntzen, E., & Holth, P. (2000). Probability of stimulus equivalence as a function of class size vs. number of classes. *The Psychological Record, 50,* 79–104. doi:https://doi.org/10.1007/BF033

Arntzen, E., & Lian, T. (2010). Trained and derived relations with pictures as nodes. *The Psychological Record, 60,* 659–677. doi:https://doi.org/10.1007/BF03395738

Arntzen, E., & Nartey, R. K. (2018). Equivalence class formation as a function of preliminary training with pictorial stimuli. *Journal of the Experimental Analysis of Behavior, 110*(2), 275–291. doi:https://doi.org/10.1002/jeab.466

Arntzen, E., Nartey, R. K., & Fields, L. (2014). Identity and delay functions of meaningful stimuli: Enhanced equivalence class formation. *The Psychological Record, 64,* 349–360. doi:https://doi.org/10.1007/s40732-014-0066-3

Arntzen, E., Nartey, R. K., & Fields, L. (2015a). Enhanced equivalence class formation by the delay and relational functions of meaningful stimuli. *Journal of the Experimental Analysis of Behavior, 103,* 524–541. doi:https://doi.org/10.1002/jeab.152

Arntzen, E., Nartey, R. K., & Fields, L. (2018a). Graded delay, enhanced equivalence class formation, and meaning. *The Psychological Record, 68,* 123–140. doi:https://doi.org/10.1007/s40732-018-0271-6

Arntzen, E., Nartey, R. K., & Fields, L. (2018b). Reorganization of equivalence classes: Effects of preliminary training and meaningful stimuli. *Journal of the Experimental Analysis of Behavior, 91,* 564–586. doi:https://doi.org/10.1002/jeab.329

Arntzen, E., & Nikolaisen, S. L. (2011). Establishing equivalence classes in children using familiar and abstract stimuli and many-to-one and one-to-many training structures. *European Journal of Behavior Analysis, 12,* 105–120. doi:https://doi.org/10.1080/15021149.2011.11434358

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

Barron, R., Leslie, J. C., & Smyth, S. (2018). Teaching real-world categories using touchscreen equivalence-based instruction. *The Psychological Record,* 89–101. doi:https://doi.org/10.1007/s40732-018-0277-0

Bentall, R. P., Dickins, D. W., & Fox, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *The Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 46B,* 187–214. doi:https://doi.org/10.1080/14640749308401085

Bortoloti, R., & de Rose, J. C. (2009). Assessment of the relatedness of equivalent stimuli through a semantic differential. *The Psychological Record, 59,* 563–590. doi: https://doi.org/10.1007/BF03395682

Bortoloti, R., & De Rose, J. C. (2012). Equivalent stimuli are more strongly related after training with delayed matching than after simultaneous matching: A study using the implicit relational assement procedure (IRAP). *The Psychological Record, 62,* 41–54. doi: https://doi.org/10.1007/BF03395785

Bortoloti, R., Rodrigues, N. C., Cortez, M. D., Pimentel, N., & de Rose, J. C. (2013). Overtraining increases the strength of equivalence relations. *Psychology & Neuroscience, 6*(3), 357–364. doi:https://doi.org/10.3922/j.psns.2013.3.13

de Almeida, J. H., & de Rose, J. C. (2015). Changing the meaningfulness of abstract stimuli by the reorganization of equivalence classes: Effects of delayed matching. *The Psychological Record, 65,* 451–461. doi: https://doi.org/10.1007/s40732-015-0120-9

Doughty, A. H., Leake, L. W., & Stoudemire, M. L. (2014). Failure to observe untested derived stimulus relations in extinction: Implications for understanding stimulus-equivalence formation. *Journal of the Experimental Analysis of Behavior, 102*(3), 311–326. doi: https://doi.org/10.1002/jeab.111

Dube, W. V., & McIlvane, W. J. (1996). Some implications of a stimulus control topography analysis. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 197–218). Amsterdam: Elsevier.

Fields, L., Adams, B. J., Verhave, T., & Newman, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior, 53,* 345–358. doi:https://doi.org/10.1901/jeab.1990.53-345

Fields, L., & Arntzen, E. (2018). Meaningful stimuli and the enhancement of equivalence class formation. *Perspectives on Behavior Science, 41,* 63–93. doi:https://doi.org10.1007/s40614-017-0134-5

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:https://doi.org/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 the Experimental Analysis of Behavior, 97,* 163–181. doi:https://doi.org/10.1901/jeab.2012.97-163

Fields, L., & Moss, P. (2007). Stimulus relatedness in equivalence classes: Interaction of nodality and contingency. *European Journal of Behavior Analysis, 8,* 141–159. doi:https://doi.org/10.1080/15021149.2007.11434279

Holth, P., & Arntzen, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record, 48,* 81–110. doi:https://doi.org/10.1007/BF03395260

Hrbacek, K., & Jech, T. (1999). *Introduction to mathematical set theory* (3rd ed.. New York, NY: Marcel Deccer.

Imam, A. A., & Blanche, J. V. (2013). Effects of a CARO on stimulus equivalence formation: A systematic replication. *The Psychological Record, 63,* 141–156. doi: https://doi.org/10.11133/j.tpr.2013.63.1.011

Jennings, A. M., & Miguel, C. F. (2017). Training intraverbal bidirectional naming to establish generalized equivalence class performances. *Journal of the Experimental Analysis of Behavior, 108*(2), 269–289. doi: https://doi.org/10.1002/jeab.277

Kato, O. M., de Rose, J. C., & Faleiros, P. B. (2008). Topography of responses in conditional discrimination influences formation of equivalence classes. *The Psychological Record, 58,* 245–267. doi:https://doi.org/10.1007/BF03395614

Kennedy, C. H. (1991). Equivalence class formation influenced by the number of nodes separating stimuli. *Behavioural Processes, 24,* 219–245. doi:https://doi.org/10.1016/0376-6357(91)90077-D

Lyddy, F., Barnes-Holmes, D., & Hampson, P. J. (2000). The effect of stimulus meaningfulness on the formation of equivalence classes. *European Journal of Behavior Analysis, 1,* 71–87. doi:https://doi.org/10.1080/15021149.2000.11434157

Mandell, C., & Sheen, V. (1994). Equivalence class formation as a function of the pronounceability of the sample stimulus. *Behavioural Processes, 32,* 29–46. doi: https://doi.org/10.1016/0376-6357(94)90025-6

Mensah, J., & Arntzen, E. (2017). Effects of meaningful stimuli contained in different numbers of classes on equivalence class formation. *The Psychological Record, 67,* 325–336. doi:https://doi.org/10.1007/s40732-016-0215-y

Moss-Lourenco, P., & Fields, L. (2011). Nodal structure and stimulus relatedness in equivalence classes: Post-class formation preference tests. *Journal of the Experimental Analysis of Behavior, 95,* 343–368. doi:https://doi.org/10.1901/jeab.2011.95-343

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

Nartey, R. K., Arntzen, E., & Fields, L. (2015a). Enhancement of equivalence class formation by pretraining discriminative functions. *Learning & Behavior, 43*(1), 20–31. doi:https://doi.org/10.3758/s13420-014-0158-6

Nartey, R. K., Arntzen, E., & Fields, L. (2015b). Training order and structural location of meaningful stimuli: Effects on equivalence class formation. *Learning & Behavior, 43,* 342–353. doi:https://doi.org/10.3758/s13420-015-0183-0.

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

O'Connor, J., Rafferty, A., Barnes-Holmes, D., & Barnes-Holmes, Y. (2009). The role of verbal behavior, stimulus nameability, and familiarity on the equivalence performances of autistic and normally developing children. *The Psychological Record, 59,* 53–74. doi: https://doi.org/10.1007/BF03395649

Pilgrim, C. (2015). Stimulus control and generalization. In F. D. D. Reed & D. D. Reed (Eds.), *Autism service delivery: Bridging the gap between science and practice* (pp. 25–74). New York, NY: Springer Science.

Randell, T., & Remington, B. (2006). Equivalence relations, contextual control, and naming. *Journal of the*

Experimental Analysis of Behavior, *86*(3), 337–354. doi: https://doi.org/10.1901/jeab.2006.82-05

Rustad Bevolden, K., & Arntzen, E. (2018). How robust is the correspondence between responding on matching-to-sample tests and sorting tests? *Norwegian Journal of Behavior Analysis*, *45*, 125–142. Retrieved from http://www.nta.atferd.no/journalissue.aspx?IdDocument=729

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

Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior*, *43*, 21–42. doi:https://doi.org/10.1901/jeab.1985.43-21

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

Smeets, P. M., & Barnes-Holmes, D. (2005). Establishing equivalence classes in preschool children with one-to-many and many-to-one training protocols. *Behavioural Processes*, *69*, 281–293. doi:https://doi.org/10.1016/j.beproc.2004.12.009

Travis, R. W., Fields, L., & Arntzen, E. (2014). Discriminative functions and over-training as class-enhancing determinants of meaningful stimuli. *Journal of the Experimental Analysis of Behavior*, *102*, 47–65. doi: https://doi.org/10.1002/jeab.91

Varelas, A., & Fields, L. (2017). Equivalence based instruction by group based clicker training and sorting tests. *The Psychological Record*, *67*(1), 71–80. doi:https://doi.org/10.1007/s40732-016-0208-x

Vie, A., & Arntzen, E. (2017). Talk-aloud protocols during conditional discrimination training and equivalence class formation. *The Analysis of Verbal Behavior*, *33*(1), 80–97. doi:https://doi.org/10.1007/s40616-017-0081-y

Received: June 7, 2018
Final Acceptance: December 18, 2019
Editor in Chief: Amy L. Odum
Associate Editor: Anna I. Petursdottir