



Children's emergent relations of equivalence using stimuli with opposite verbal labels: Exclusion and minimal training conditions

Jacqueline J. Schenk^{a,*}, Mickey Keenan^b, Harrie H. Boelens^c, Simon Dymond^{d,e}, Paul M. Smeets^c

^a Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, the Netherlands

^b University of Ulster, Department of Psychology, Coleraine, N-Ireland, United Kingdom

^c Department of Psychology, Leiden University, the Netherlands

^d Department of Psychology, Swansea University, United Kingdom

^e Department of Psychology, Reykjavík University, Menntavegur 1, Nauthólsvík, 101 Reykjavík, Iceland

ARTICLE INFO

Keywords:

Matching-to-sample
Responding by exclusion
Stimulus equivalence
Emergent relations
Children

ABSTRACT

The present study examined different conditions under which exclusion responding in conditional discrimination tasks would generate emergent equivalence relations in young children based on shared relationships with verbal labels. Both visual stimuli (Sets A, B, C, and D) and auditory stimuli (spoken words, Set N: N1 “correct”; N2: “incorrect”) were used. Following a pilot study, three experiments were conducted, each involving eight preschool children. These experiments systematically investigated under which conditions responding by exclusion (i.e., responding away from a designated S^- comparison in a matching to sample context) would generate sufficiently stable sample- S^+ relations for arbitrary stimulus classes to establish. The results showed that young children's exclusion responding under test conditions will only contribute to arbitrary stimulus class formation and expansion when training has already established two arbitrary stimulus classes involving at least two stimuli each. For young children to demonstrate emergent conditional discrimination performances that are indicative of the formation of equivalence relations, it is necessary to have training and/or reinforced exposure to both S^+ and S^- control elements required for deriving the appropriate emergent relations with at least two conditional relations involving different samples. These findings not only contribute to existing research and theory on the conditions under which exclusion responding may contribute to fundamental language and learning processes, they also contribute to the experimental predictability of emergent conditional matching behaviours in preschool children by further unravelling the conditions under which emergent matching based on exclusion generates arbitrary conditional relations of equivalence.

Experimental developmental psychologists have long been interested in how young children come to show behaviour that is new and cannot be explained in terms of an explicit history of reinforcement (e.g., Barnes et al., 1995; Schenk, 1995; Smeets et al., 1996). Conditional discrimination procedures are often used in these studies. For example, in an arbitrary (or symbolic) matching-to-sample (MTS) procedure two or more stimuli, the samples, are successively presented. In the presence of each sample stimulus, at least two other stimuli, the comparisons, are provided in order for a choice to be made between them. For example, a participant is taught to select comparison B1 (and not B2) in the presence of sample A1, and to select comparison B2 (and not B1) in the

presence of sample A2 (Cumming and Berryman, 1965; McIlvane et al., 1987; Sidman et al., 1982; Sidman and Tailby, 1982).

After mastering arbitrary conditional discriminations A–B and B–C, participants often demonstrate new conditional relations that are derived from the previously trained relations (e.g., Sidman, 1994, 2000). Typically, these performances exhibit the three properties of equivalence: reflexivity, symmetry, and transitivity (Sidman, 1992). To test for these properties, participants are exposed to various probes including A–A, B–B, and C–C probes (reflexivity), B–A and C–B probes (symmetry), A–C probes (transitivity), and C–A probes (combined symmetry/transitivity). When these relations are demonstrated, it can be

* Corresponding author at: Erasmus University Rotterdam, Erasmus School of Social and Behavioural Sciences, Burgemeester Oudlaan 50, Mandeville Building T13-6, 3062 PA Rotterdam, the Netherlands.

E-mail address: schenk@essb.eur.nl (J.J. Schenk).

<https://doi.org/10.1016/j.beproc.2021.104341>

concluded that relations between the A, B, and C stimuli constitute ‘equivalence relations’, and as such they evidence the presence of two classes of equivalent stimuli, A1B1C1 and A2B2C2.

Conditional discriminations described above between A and B stimuli implies the rule “If A1, then B1” and “If A2, then B2.” A sample-comparison relation such as A1-B1, may be based on the participant selecting comparison B1 with sample A1 (S^+ control) and/or rejecting comparison B2 with sample A1 (S^- control) (for a detailed theoretical analysis of control by negative stimuli, see Carrigan and Sidman, 1992). A more general form of S^- control is called control by ‘exclusion’. (Dixon et al., 1983; Tomonaga, 1993).

A typical example of exclusion responding would be the following scenario: a participant first learns to select comparison B1 (and not B2) with sample A1, and select comparison B2 (and not B1) with sample A2. Thereafter, s/he is presented with sample A1 but now with Y1 and B2 as comparisons, and sample A2 with Y2 and B1 as comparisons. Based on previous studies, it is expected that with sample A1, this participant selects Y1 (and not B2) and with sample A2, selects Y2 (and not B1). When these responses are observed without any additional training, then, apparently, sample A1 not only controls the selection of the S^+ but also controls the rejection of S^- . Similarly, sample A2 not only controls the selection of B2 but also the rejection of B1. These controlling ‘select’ and ‘reject’ relations (Kato et al., 2008) or Relation S and Relation R (Carrigan and Sidman, 1992) may not only generate accurate responding on conditional discrimination training and test trials but also yield accurate responding on equivalence tests. The study of responding by exclusion is considered important because it is generally accepted that, in addition to positive control, negative control can also function as a source of information for participant’s selection comparisons (e.g. Johnson and Sidman, 1993). Compared to positive control, however, the study of negative control in discrimination learning, is relatively understudied and deserves more experimentation considering its promising potential to be effectively used to efficiently teach young learners arbitrary conditional discriminations and/or to yield equivalence performances in applied settings (e.g. Langsdorff et al., 2017; Plazas and Villamil, 2018).

Concepts that are closely related to responding by exclusion are ‘fast mapping’ (Carey and Bartlett, 1978) or ‘emergent mapping’ (Wilkinson et al., 1996). These concepts refer to participants to, accurately, immediately, and often fast matching a novel word to a novel object in the presence of familiar (i.e. previously defined) objects. These concepts are often studied in psycholinguistic studies (e.g., Mervis and Bertrand, 1994) to explain (a) the rapid quantitative expansion of children’s (expressive and receptive) vocabulary (Ganger and Brent, 2004; Nazzi and Bertoncini, 2003), and (b) to study the application of exclusion responding when teaching, e.g., language, reading, and spelling skills (e.g., Costa et al., 2001; Dixon, 1977; Grassmann et al., 2009, 2015; McIlvane et al., 1984; Langsdorff et al., 2017; McIlvane et al., 1992b; Stromer and Osborne, 1982; Wilkinson et al., 2009). In addition, exclusion responding is also of interest to experimental and applied behaviour-analytic studies (e.g., Dixon, 1977; Lipkens et al., 1993; McIlvane et al., 1992a,1992b; McIlvane et al., 1988; McIlvane and Stoddard, 1981; McIlvane et al., 1984; Wilkinson and Green, 1998), particularly for (1) the study of the conditions under which responding by exclusion reliably predicts the emergence of stable arbitrary stimulus-stimulus relations (or not), and (2) for the study of stimulus equivalence (Carr and Felce, 2000).

Exclusion responding can be observed with a range of stimuli and a range of non-human populations, such as dogs (Aust et al., 2008; Zaine et al., 2016), rats (e.g., de Souza and Schmidt, 2014), and chimpanzees (e.g., Beran, and Washburn, 2002; Tomonaga, 1993) as well as human populations that vary in chronological and/or mental age (Bates, 1979; Cippola et al., 2014; Costa et al., 2013, 2001; Dixon, 1977; Domeniconi et al., 2007; Grassmann et al., 2009, 2015; Langsdorff et al., 2017; Meehan, 1995; McIlvane et al., 1984, 1992a,1992b; McIlvane et al., 1988; Plazas and Villamil, 2018; Roberta et al., 2001; Stromer and

Osborne, 1982; Wilkinson et al., 2000, 2009). In the study by Costa et al. (2013), for example, eight 4–5-year-old pre-schoolers were taught three relations between pictures (the comparisons) and dictated words (samples) using matching-to-sample procedures. Thereafter, a novel, undefined dictated word was presented as a sample, and three comparisons were presented, two of which were defined, and one was undefined. On these so-called control trials, the undefined comparisons were either a blank comparison or a novel comparison. Subsequently, exclusion trials were interspersed among baseline trials, on which undefined dictated words served as sample stimuli and an undefined picture, a defined picture and a blank comparison served as comparison stimuli. Selecting the undefined picture with an undefined dictated word as sample was considered responding by exclusion. The results of this study showed that all participants responded by exclusion on all trials. Although the performances of the majority of participants showed (a) accurate naming of the undefined pictures and (2) consistent responding on MTS probes based on sample- S^+ and sample- S^- relations, these performances emerged gradually (after being exposed to 3–10 exclusion trials). The authors concluded, in line with earlier studies (e.g., Costa et al., 2001; Domeniconi et al., 2017; Wilkinson and McIlvane, 1997), that learning (i.e., accurately and reliably showing sample- S^+ relations) after being exposed to exclusion trials is more likely to emerge with typically developing children, provided that the range of prerequisite behaviours (i.e., repertoire) of the participants and aspects of the teaching procedure (e.g., contingency arrangements and exposure to a minimum number of exclusion trials) are taken into account. Regarding these exclusion trials, one of the recommendations from studies on the application of exclusion responding, however, is that exclusion trials need to be interspersed with well-established baseline trials for effective teaching (de Souza et al., 2009). This recommendation is based on the fact that relations that are established by exclusion responding may be of concern for further learning and therefore sufficient opportunity (i.e., number of trials) needs to be given to participants (McIlvane et al., 1984; Wilkinson et al., 2009) before proceeding to further learning or testing (Costa et al., 2001; Domeniconi et al., 2007; Wilkinson and McIlvane, 1997). Other recommendations to encourage the establishment of stable sample- S^+ relations during or after responding on exclusion trials, include the use of differential consequences contingently following upon selection by exclusion (e.g., Carr, 2003; Ferrari et al., 1993, 2008; McIlvane and Stoddard, 1981) or the training procedure (Langsdorff et al., 2015), the use of manipulable stimulus materials with young infants as participants (Domeniconi et al., 2007), exposing participants to more than one exclusion responding opportunities (number of trials) (Costa et al., 2001; Domeniconi et al., 2007; McIlvane et al., 2009; Wilkinson and McIlvane, 1997), a gradual introduction of new undefined stimuli or relations on exclusion probes, only proceeding to the next relation after tests of performance show the participant to have mastered the previous relation (McIlvane et al., 1992a,1992b); see Langsdorff et al. (2017) for an overview of recommendations for procedures that aim to achieve stable exclusion responding when teaching auditory-visual relations. The current study further examined the effects of contingency manipulation on emergent responding by exclusion and subsequent effects on the stability and equivalence properties of these conditional relations (Carr, 2003; Ferrari et al., 1993, 2008; McIlvane and Stoddard, 1981).

The present study included three experiments that further examined preschool children’s emergent arbitrary matching performances via exclusion under controlled, two-choice conditions. A pilot study that preceded the currently reported experiments demonstrated that normally developing preschool children are very capable of demonstrating emergent equivalence relations between stimuli that were both associated with the spoken word “correct” (N1) and stimuli that were both associated with the spoken word “incorrect” (N2) (see *Supplementary materials*: Supplementary Fig. 1 for the design of the pilot study). In short, children were trained to conditionally select A1 (S^+) when asked to select the “correct” comparison and to select A2 (S^+) when asked to

select the “incorrect” comparison (N–A conditional discrimination training). Similarly, they also received N–B training and A–A identity matching training. On subsequent A–B and B–A probes, all eight children conditionally related A1–B1 and A2–B2 and vice versa. These results (see Supplementary Tables 1 and 2: *Supplementary Materials*) showed that equivalence classes can be established based on shared relations with verbal labels with opposite meanings (N1: “correct”) and N2: “incorrect”). The relevance of the pilot study was to show that arbitrary stimulus classes can be established through shared relations with not only “correct” but also “incorrect” verbal labels, thereby bridging the previously described simple discrimination with the typical MTS format.

To date, exclusion research has focused on several procedures under which exclusion responding can be evoked and may lead to stable demonstration of sample-S⁺ relations. The primary aim of the current study, however, is to examine the minimal contingency arrangements by which sample-comparison relations that are established under exclusion conditions would generate arbitrary stimulus equivalence classes.

1. Experiment 1

The goal of Experiment 1 was to first examine whether N2–A2 and N2–B2 matching performances can be established via exclusion responding after N1–A1 and N1–B1 training only. Based on the findings of other studies on exclusion responding with preschool children as participants (e.g., Costa et al., 2013; Dixon and Dixon, 1978; Costa et al., 2001; McIlvane et al., 1988), it was expected that after N1–A1 and N1–B1 relations are established, N2–A2 and N2–B2 would emerge

during testing via exclusion responding. Additionally, Experiment 1 investigated whether the established relations N1–A1 and N1–B1, and the emergent relations N2–A2 and N2–B2, would also generate emergent conditional relations between A and B stimuli as observed in the pilot study (A–B and B–A probes). If so, this would be indicative of the formation A1B1N1 and A2B2N2 arbitrary stimulus classes (for design details of Experiment 1, see Fig. 1, top row).

2. Method

2.1. Participants, setting, and sessions

Two groups of four children participated. Group 1 (Participants 1–4) consisted of two girls and two boys aged between 5,3 and 5,8 years (mean age: 5,5 years). Group 2 (Participants 5–8) consisted of three girls and one boy aged between 5 and 5,7 years (mean age: 5,3 years).

Sessions were conducted individually in a quiet room of the school building, five times a week, once or twice a day. Sessions consisted of 12–36 trials and lasted for 4–7 min each day. The children were recruited from an elementary school and had never participated in similar experiments. Active parental consent was obtained for each participant. One adult served as the experimenter who sat facing the child across a table. Prior to her participation to this study, the experimenter had received extensive training in the prevention of subtle cues (facial expression, eye darting) that could influence the participants’ responses (see Dymond et al., 2005). Another adult served as reliability observer. When reliability checks were made, the observer was present in the same room and located such that the participant’s responses could

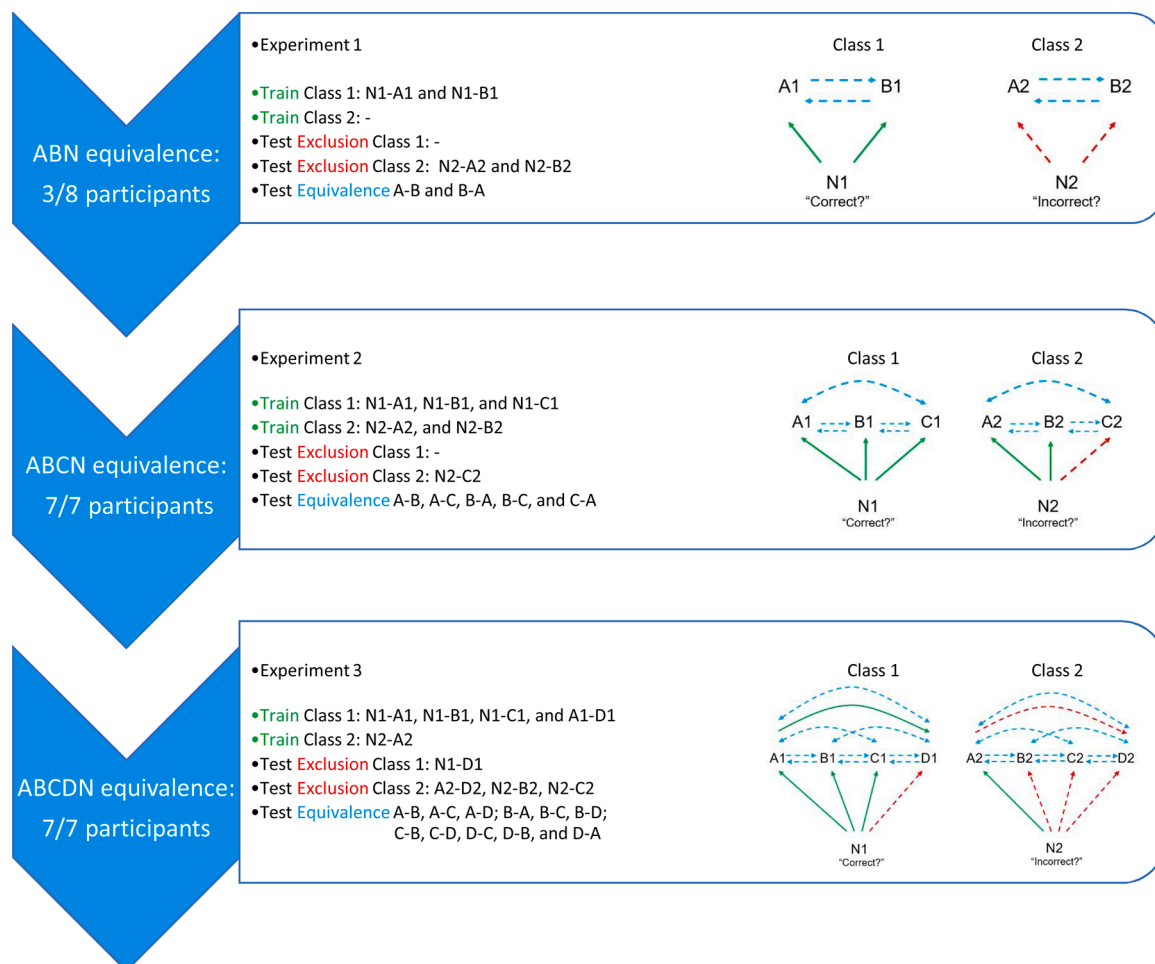


Fig. 1. xxx.

be clearly observed. In all sessions, the experimenter and observer (if present) used a previously prepared data sheet to record the participant's selections on every trial. The observer could not observe the experimenter's recordings. Sessions were conducted five times a week, once or twice a day. Reliability scores were made for 58 % of the sessions.

2.2. Materials

Stimuli that were used in Experiment 1 were the same as those used in the pilot study and involved arbitrary visual stimuli (Sets A and B) and verbal stimuli (Set N). Verbal stimuli involved two questions asked by the experimenter: "Which one is correct?" (N1) and "Which one is incorrect?" (N2). Visual stimuli were two coloured squares, one coloured red (A1) and one coloured blue (A2), and two shapes, a bow (B1) and a merlon-shaped form (B2) that were drawn in black ink. Fig. 2 shows the stimuli used on training and test trials.

Stimuli were presented on white cards that measured 14.7 × 10.5 cm. On these cards, either two comparison stimuli, or two comparison stimuli and one sample stimulus were presented. A and B visual stimuli filled an area of at most 1.5 × 1.5 cm. Comparison stimuli were located at 7 cm of each other and, if a sample stimulus was involved, located at 3.2 cm of the sample stimulus (spacings centre to centre). All sample stimuli were presented equally often. Each comparison stimulus appeared in the left and right position and as the correct (S⁺) and incorrect (S⁻) comparison an equal number of times. The position of the S⁺ was varied unsystematically, with the restriction that neither position was correct more than three times consecutively. Additional materials were a little vase (20 cm in height) made of clear glass and a tray containing wooden beads. The vase showed a horizontal mark that would be reached when the vase contained 40 beads.

2.3. Procedure

2.3.1. Instructions and reinforcement procedures

At the beginning of the experiment, the experimenter put eight pictures (cartoon characters, dinosaurs, etc.) on the table and told the child: "We are going to play a game and if you play it well you can earn a picture. Which one do you like best?" When the child selected one of the pictures, the experimenter removed the other pictures from the table, placed the picture selected at the corner of the table, and placed the vase and the beads-containing tray on the table. The child was told: "If you play it well, I will say: Yes! And you may take a bead. You can put the bead in here [the experimenter points to the horizontal mark on top of the vase] and if the tray is empty you have earned your picture. If you don't play it well, I will say: "No! And you are not allowed to take a bead."

Training sessions consisted of two blocks of 12 training trials or one block of 24 training trials. On training trials, every correct response was followed by the delivery of a bead. Training involved trials on which two visual stimuli (comparison stimuli) were presented on a card and trials (identity MTS) on which three visual stimuli (a sample and two comparisons) were presented on a card. When a card was presented with two stimuli, the experimenter asked the child: "Which one is correct?" or "Which one is incorrect?" Throughout the manuscript, N1 refers to the question "Which one is correct?" and N2 refers to the question "Which one is incorrect?" If these questions are asked in the presence of, for example, stimuli X1 and X2 then training is denoted as N-X. During training, N1 and N2 were presented equally often. On identity MTS trials, the experimenter pointed to the sample and asked the child: "Which one goes with this one?" Pointing to the designated accurate comparison was followed by the experimenter saying "Yes! You may take a bead." Pointing to the designated inaccurate comparison was followed by the experimenter saying "No! You may take no bead." Invalid responses such as pointing to both stimuli were followed by a corrective remark, e.g., "You should look at the pictures when pointing."

Test trials were conducted without consequences apart from the

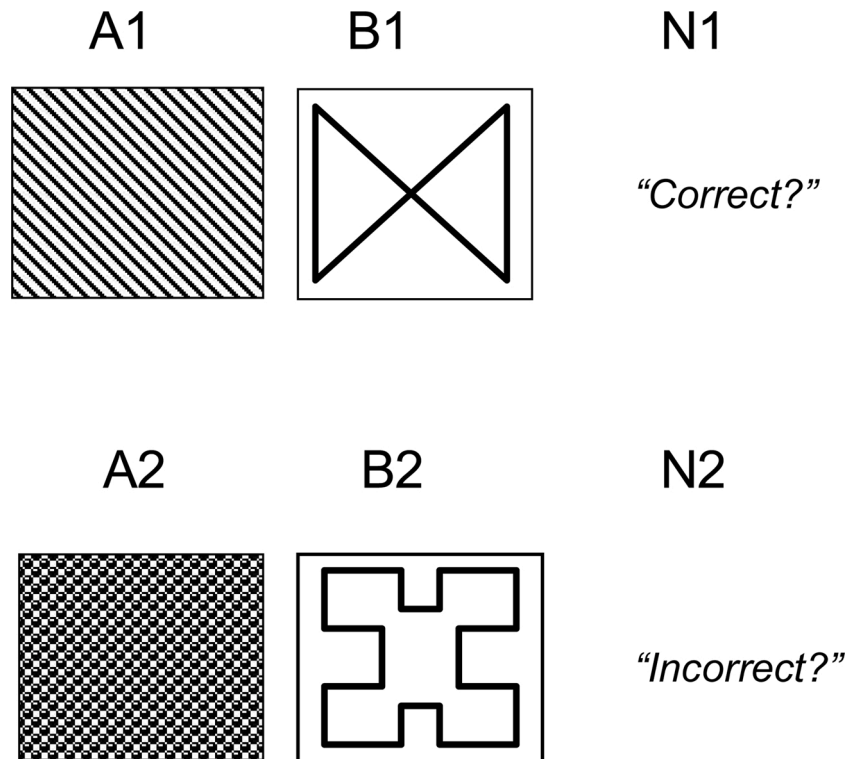


Fig. 2. Stimulus series A, B and N used in the pilot study and Experiment 1. Colours are represented by patterns: A1 = Red square; A2 = Blue square.

removal of the stimulus card. In test sessions, training trials were the same as in training sessions. At the beginning of each block of test trials, the experimenter removed the tray that contained the beads and the vase from the table while saying, "Now we are going to play the game without me saying if you played it well or not. You cannot earn any beads now. Do the best you can." On test trials involving N1 and/or N2 as sample stimuli, the instructions were the same as for training trials involving N1 and/or N2 as samples. On A-A identity matching and A-B and B-A arbitrary-matching test trials, no instructions were given. After the participant selected a comparison, the experimenter removed the card from the table without any further comment and went on to the next trial.

Throughout training and testing, stringent criteria are used during each phase of each experiment that seek to ensure firm establishment of not only baseline conditional relations, but also unambiguous demonstration of emergent performances on test trials. This is rather important for MTS tasks involving two choices only (Boelens, 2002).

2.4. Training and Tested Relations

Fig. 1 (top row) illustrates the relations that were trained in Experiment 1 (solid arrows). Baseline training consisted of N1-A1 and N1-B1 training trials, and A-A identity-matching tasks. N1-A1 was trained in two blocks of 12 trials (three blocks maximum). Mastery criterion required correct responding on at least 11/12 trials in the second block. N1-B1 was trained in the same way as N1-A1. Thereafter, participants received one block of six N1-A1 trials randomly mixed with six N1-B1 trials. Criterion required correct responding on at least 11/12 trials.

A-A identity-matching training

Participants received blocks of 12 A-A identity-matching training trials in order to familiarise them with a matching-to-sample format with visual stimuli only. Selections of the comparison identical to the sample were reinforced. Criterion was reached if correct responses were made on at least 11/12 trials in a block

Maintenance testing

During maintenance testing, participants received two blocks of six test trials each (four A-A, one N1-A1, one N1-B1 trial) and two blocks of six training trials each (four A-A, one N1-A1, and one N1-B1 trial). This test was conducted to verify that criterion performances were maintained under non-reinforcement conditions. The criterion required correct responding on at least 11/12 test trials.

N-A arbitrary matching probes

To examine the emergence of N2-A2 via exclusion responding (i.e., excluding A1 when N2 is the sample), participants received three blocks (Blocks 1, 3, and 5) of 12 test trials each (eight N2-A2, two N1-A1, and two A-A trials). They also received three blocks (Blocks 2, 4, and 6) of four training trials each (two A-A, one N1-A1, and one N1-B1 trial). As recommended by several studies (e.g. de Souza et al., 2009), exclusion probes were interspersed among other probes. Criterion required accurate responding on at least 22/24 N2-A2 trials, 11/12 maintenance trials, and 11/12 baseline-training trials.

N-B arbitrary matching probes

The emergence of N2-B2 was tested in the same way as N2-A2, except that in each test block eight N2-B2 probe trials were mixed with two N1-B1 and two A-A trials.

A-B and B-A arbitrary matching probes

This test assessed emergent conditional relations between B and A stimuli. The samples were B1 or B2, and the comparisons were A1 and A2. Participants received two blocks of ten test trials; each block consisted of four B-A trials that were irregularly mixed with two N-A trials, two N-B trials, two A-A trials. Participants also received two blocks of eight training trials each (see Maintenance testing). Criterion required correct responding on at least 7/8 B-A probes, 11/12 maintenance trials and 15/16 training trials. This test was repeated once if criterion was not reached in the first session.

3. Results and discussion

The experiment consisted of 1576 trials, 1088 training and 488 test trials. Reliability scores were made on 928 trials (58.9 %), on 640 training trials (58.8 %) and on 288 test trials (59 %). The experimenter's recordings agreed with the observer's recordings on all trials.

All participants reached the criteria during both N1-A1 and N1-B1 training in the second block of 12 training trials. When the N1-A1 and N1-B1 trials were mixed on 12 trials, all eight participants responded at criterion level on the first session. During A-A training, all participants responded perfectly on the first 12-trial block.

On all test sessions, A-A identity matching, and N1-A1 and N1-B1 matching performances were maintained by participants. On maintenance trials, all participants responded at criterion level on both test and training trials. Table 1 presents the results on N2-A2 and N2-B2 probes. On sessions with N2-A2 probes trials, Participants 3, 4, 6, and 7 selected comparison A2 on all 24 N2-A2 probes. Participants 1, 2, and 8 selected A1 on the first N2-A2 probe trial and A2 on the 23 remaining N2-A2 trials. Participant 5 selected A1 on all 24 N2-A2 test trials. On N2-B2 test sessions, all participants except Participant 5 selected B2 on all 24 N2-B2 probe trials; Participant 5 selected B1 on all N2-B2 probe trials. When both N2-A2 and N2-B2 test sessions were repeated, however, Participant 5 selected A2 on 23/24 N2-A2 trials and B2 on 24/24 N2-B2 probe trials.

Table 2 shows the results on the B-A and A-B probes (results on B-A scores are the added scores on eight B-A probes during B-A testing and four B-A probes during A-B testing). On B-A probes, only Participant 4 demonstrated emergent conditional relations between B and A stimuli on 12/12 B-A trials. Participants 1, 2, 3, 5, 7, and 8 selected A1 on most B-A probes (range: 50–58 % accurate matching). Participant 6 made accurate selections on 10/12 B-A trials. During A-B testing, Participants 4, 5, and 6 reliably showed the expected conditional relations between A and B stimuli; Participants 1, 2, and 3 selected B1 on all trials while Participants 7 and 8 selected B2 on most trials (range: 50–66.7 % accurate matching). Thereafter, all participants except Participant 4 received a second B-A test session in which Participants 1, 2, 3, 7, and 8 mainly selected A1 (range: 50–58 % accurate matching). Participant 5 responded accurately on 11/12 B-A probes and Participant 6 on 12/12 B-A probes. Finally, Participants 1, 2, 3, 7, and 8 received a second A-B test session in which Participants 1, 2, and 3 mainly selected B1 and Participants 7 and 8 mainly selected B2 (range: 41.7–58.3 % accurate matching). Thus, only three participants (Participants 4, 5, and 6) demonstrated the expected B-A and A-B matching performances.

The results of Experiment 1 may be summarised as follows: (1) all children reliably demonstrated N1-A1 and N1-B1 matching and A-A identity matching task; (2) all participants except Participant 5 immediately showed N2-A2 and N2-B2 matching via exclusion; Participant 5 showed these performances when the test sessions involved were repeated; (3) despite showing N-A and N-B conditional discriminations, only three of eight participants (eventually) showed the expected emergent conditional discrimination performances on B-A and A-B arbitrary-matching probes; the remaining five participants responded around chance level.

Table 1

Experiment 1: Individual test scores on N2-A2 and N2-B2 probes. The results are presented as the number of test trials on which responding was consistent with stimulus class formation/the total number of probe trials.

Participant no.								
	1	2	3	4	5	6	7	8
Probe								
N2-A2	23/24	23/24	24/24	24/24	0/24;	24/24	24/24	23/24
N2-B2	24/24	24/24	24/24	24/24	0/24;	24/24	24/24	24/24

Table 2

Experiment 1: Individual test scores on arbitrary matching probes. The results are presented as the number of test trials on which responding was consistent with stimulus class formation/the total number of probe trials of each type during B-A and A-B testing.

Test	Participant no.							
	1	2	3	4	5	6	7	8
B-A	6/12	7/12	7/12	12/12	6/12	10/12	6/12	6/12
A-B	6/12	6/12	6/12	11/12	12/12	12/12	6/12	8/12
B-A	6/12	6/12	7/12	–	11/12	12/12	6/12	6/12
A-B	5/12	6/12	6/12	–	–	–	6/12	7/12

Note. "–" = The participant did not receive these probes.

With regard to N2-A2 and N2-B2 testing, the results are consistent with previous findings on exclusion responding in arbitrary-matching tasks with children participating (e.g., Costa et al., 2001, 2013; Dixon and Dixon, 1978; McIlvane et al., 1988; Stromer, 1986, 1989). Although in the current study, all participants demonstrated N1-A1 and N1-B1 matching during training and maintenance testing and N2-A2 and N2-B2 matching during testing, only a minority of children subsequently showed emergent B-A and A-B matching performances. The number of occasions (i.e., probes) on which participants could match N2-A2 and N2-B2 matched the number of occasions on which participants were taught to match N2-A2 and N2-B2 during training in the pilot study, thus, the amount of exposure per se does not seem to be the explaining factor for participants in the pilot study to be able to show transitive and equivalence relations between A-B and B-A. The critical element was more about the exposure to reinforced examples of negative stimulus control within the context of conditional discriminations, rather than exposure to additional S⁺ control defined by the reinforcement contingencies between the sample and the comparison stimuli. It might even be the combination of both that is critical. In sum, although the children in Experiment 1 demonstrated emergence of auditory-visual relations through exclusion, these demonstrations of exclusion relations were never reinforced. Therefore, there might be an insufficiently strong basis for further stimulus matching through S⁻ control in the absence of reinforcement of the exclusion-derived relations.

4. Experiment 2

In Experiment 1, an analysis of the results on B-A and A-B probes for participants who failed to show emergent matching performances reveals that these participants mainly selected those A and B stimuli, respectively, to which responding was reinforced during N1-A1 and N1-B1 (or N1-B2) training: Participants 1, 2, and 3 mainly selected A1 and B1, and Participants 7 and 8 selected A1 and B2 on most B-A and A-B probes, respectively. It might be suggested that the procedure has taught children that responding in the presence of N1 is "right" and responding in the presence of N2 is "wrong". After all, accurate responding in the presence of N1 was rewarded during training but accurate responding in the presence of N2 was not since this was conducted during testing only. As a result, children might simply continue to select the A and B stimuli that previously functioned as S⁺ stimuli, regardless of which sample was present. The finding that participants treated A-B and B-A matching tasks as simple discrimination tasks (i.e., selecting comparison stimuli non-conditionally) was also reported in other studies on emergent discriminations (e.g., Smeets et al., 1996), although in the pilot study, and both Experiments 1 and 2, conditional control was demonstrated by N1 and N2 (N1-A1, N1-B1 during training and maintenance, N2-A2, and N2-B2 on probes). However, comparable to, e.g., Experiment 1 in the study by Smeets and Barnes (1997), demonstrations of responding by exclusion were never reinforced. Therefore, there seems to be no strong basis for further stimulus matching through S⁻ control in the absence of reinforcement of the exclusion-derived relations. This issue was further

addressed in the following experiment by having participant's accurate responding by exclusion followed by reinforcement. Therefore, in Experiment 2, not only N1 but also N2 was involved in baseline training as in the pilot study.

5. Method

Two groups of four children participated. Group 1 (Participants 1–4) consisted of two girls and two boys between 5,1–5,4 years (mean age: 5,3 years). Group 2 (Participants 5–8) consisted of two girls and two boys between 5,1–5,4 years (mean age: 5,2 years). Setting and sessions were the same as in Experiment 1, except that another professionally trained experimenter and observer conducted the experiment. Instructions and reinforcement procedures were the same as in Experiment 1. Reliability scores were obtained for about 40 % of the sessions. Stimuli that were used in Experiment 2 are shown in Fig. 3. Other materials were the same as in Experiments 1.

Fig. 1 (mid row) illustrates the relations that were trained (solid arrows) and tested (broken arrows) in Experiment 2. First, participants received baseline training, which consisted of N-A, N-B, and A-A matching tasks. Baseline tasks were trained in the same way as in Experiment 1 with the same mastery criteria. When participants maintained baseline performances under testing conditions, they received eight B-A and eight A-B probes that tested for emergent arbitrary-matching performances between A and B stimuli on the basis of shared relations with N1 and N2 respectively. Second, Experiment 2 examined the emergence of arbitrary conditional relations between novel C stimuli and the A and B stimuli. Following B-A and A-B testing, participants received N1-C1 training. Participants received blocks of 12 training trials. During N1-C1 training, visual stimuli C1 and C2 were simultaneously presented on each trial while the experimenter asked the child: "Which one is correct?" (N1). For Group 1 participants, C1 was considered correct; for Group 2 participants, C2 was correct. Reaching criterion required 11/12 correct responses within one block (maximum three blocks). Thus, participants did not receive N2-C2 (Group 1) or N2-C1 (Group 2) training trials.

After mastering N1-C1, participants received N2-C2 testing. Sessions consisted of two blocks of eight test trials each (six N2-C2 and two N1-C1 trials) and two blocks of ten training trials each (four A-A, two N-A, two N-B and 2 N1-C1 trials). Criterion required correct responding on at least 11/12 N2-C2 trials, 2/2 N1-C1 test trials, and 19/20 training trials. Then participants received eight B-C probes (Block 1) and eight C-B probes (Block 3). B-C and C-B probes were mixed with two N-B probes and two N-C probes. On N-B and on N-C probes, N1 and N2 each appeared on one trial of each probe type. Group 1 participants were expected to relate B1-C1, C1-B1 based on N1, and B2-C2, C2-B2 based on N2. Group 2 participants were expected to relate B2-C2 and C2-B2 based on N1, and B1-C1 and C1-B1 based on N2. After B-C and C-B testing, participants received eight A-C (Block 1) and eight C-A probes (Block 2). A-C and C-A probes were mixed with two N-A probes and two N-C probes. On N-A and on N-C probes, N1 and N2 each appeared on one trial of each probe type. Group 1 participants were expected to relate A1-C1, C1-A1 based on N1, and A2-C2, C2-A2 based on N2. Group 2 participants were expected to relate A1-C2 and C2-A1 based on N1, and A2-C1 and C1-A2 based on N2. These probes examined whether participants were able to demonstrate emergent relations between, for example, C1 and A1 and B1, and between C2 and A2 and B2.

6. Results and discussion

The experiment consisted of 2232 trials; 1436 training and 796 test trials. Reliability scores were made on 964 trials (43.2 %); 652 training trials (45.4 %) and 312 test trials (39.2 %). The experimenter's recordings disagreed with the observer's recordings on one test trial (N2-A2).

Participants took two to three blocks of 12 training trials to reach the

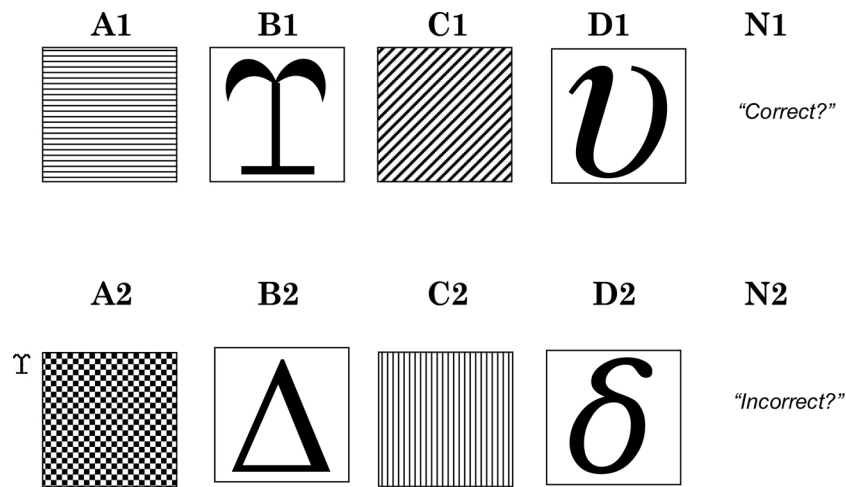


Fig. 3. Stimulus series A, B, C, and N used in Experiments 2 and 3. D stimuli were used in Experiment 3 only. Colours are represented by patterns: A1 = Pink square; A2 = Blue square; C1 = Yellow square; C2 = Green square. B and D stimuli are in black ink.

mastery criterion of the N-A task. During N-B training, all participants except Participant 6 reached criterion in the second block. Participant 6 responded correctly on 8/12 trials in the first block, on 7/12 trials in the second block, and on 6/12 trials in the third block. Therefore, Participant 6 was excluded from further participation. When N-A and N-B tasks were mixed, the remaining seven participants took one to two sessions to respond at criterion level. During A-A training, six participants responded perfectly on the first 12-trial block; only Participant 4 responded incorrectly on the first trial.

During maintenance testing, the remaining seven participants responded at criterion level at the first test session. On all subsequent test sessions, participants maintained criterion responding on baseline training and test trials. Table 3 shows the results on arbitrary-matching A-B and B-A probes. During B-A testing, all 7 remaining participants demonstrated emergent conditional relations between B and A stimuli on at least 7/8 B-A trials, except Participant 7, whose responding was class consistent on 5/8 B-A trials. Subsequently, during A-B testing, all seven participants reliably showed the expected conditional relations between A and B stimuli. Participant 7 received a second B-A test, in which she responded perfectly.

During N1-C1 training, participants reached criterion within one block of 12 trials. On N2-C2 probe trials, all seven participants selected C2 upon N2 on at least 11/12 probe trials, while responding at criterion on N1-C1 test trials and baseline training trials. On arbitrary-matching probes involving C1 and C2 (C-B, B-C, A-C, and C-A), all participants

except Participant 3 (correct responding on 5/8 probes) showed the expected matching performances. When Participant 3 received a second block involving B-C probes, this participant responded accurately on all probes.

The results of Experiment 2 can be summarised as follows: first, on B-A and A-B arbitrary-matching probes, the results for all participants who learned baseline tasks were indicative of the formation of ABN stimulus classes, results that are consistent with the results of the pilot study. Second, after participants learned N1-C1 matching during training, they subsequently matched N2-C2 during testing, results that are in accordance with the results on the exclusion probes in Experiment 1. Third, on probes that involved C1 and C2 (B-C, C-B, A-C, and C-A probes), responding for all participants who previously demonstrated B-A and A-B matching indicated that the ABN stimulus classes were expanded with C1 and C2. These results showed that all seven participants related not only those A, B, and C stimuli that shared relations with N1, but they also related those stimuli that shared relations with N2, that were either established via training (N2-A2 and N2-B2, for example), or emerged during testing (N2-C2, for example). The results, then, provide evidence for the formation of two stimulus classes with four members each: A1B1C1N1 and A2B2C2N2 for Group 1 participants, and A1B2C2N1 and A2B1C1N2 for Group 2 participants.

As for the arbitrary-matching performances involving A, B, and C stimuli, it may be suggested that the N-A, N-B, and N-C relations have symmetrical and transitive properties and that the established stimulus classes may be interpreted as stimulus equivalence classes. Experiment 3 was designed to investigate whether evidence could be found to support this suggestion.

7. Experiment 3

Based on the encouraging results on arbitrary-matching probes involving C2 in Experiment 2, it is suggested that if an undefined comparison stimulus (e.g., C2) is matched to a defined sample via exclusion responding, the comparison may become a member of the stimulus class with its respective sample. This raises the question of whether figure

(Study question 1). After all, training of N1-A1, and N2-A2 matching might already generate two AN stimulus classes, A1N1 and A2N2. Furthermore, if participants also receive training of N1-B1 matching they should be capable of demonstrating not only N2-B2 matching via exclusion but also of B-A and A-B matching performances during testing (Study question 2). In addition to finding evidence for equivalence properties of the arbitrary stimulus relations, Experiment 3 also examined whether the same results can be obtained as in Experiment 2 with regard to the formation of ABCN arbitrary stimulus classes when

Table 3

Experiment 2: Individual results on exclusion and arbitrary-matching probes of one or two sessions in which baseline responding was maintained. The results are presented as the number of trials on which the designated comparison was selected/the total number of test trials.

Participant no.	Participant no.							
	1	2	3	4	5	6	7	8
Probes								
B-A	7/8	8/8	7/8	7/8	8/8	-	5/8; 8/8	8/8
A-B	8/8	8/8	7/8	8/8	8/8	-	8/8	7/8
N2-C2	11/12	12/12	12/12	12/12	11/12	-	12/12	12/12
B-C	7/8	8/8	5/8; 8/8	8/8	8/8	-	8/8	8/8
C-B	7/8	8/8	8/8	8/8	8/8	-	8/8	7/8
A-C	8/8	8/8	7/8	8/8	8/8	-	8/8	8/8
C-A	8/8	8/8	8/8	7/8	8/8	-	8/8	7/8

Note. "-" = The participant did not receive these probes.

training involved A-A, N1-A1, N2-A2, N1-B1, and N1-C1 matching (Study question 3).

As discussed in Experiment 1, the use of spoken words N1 and N2 might have facilitated exclusion responding. Consequently, in Experiment 3, it was investigated whether participants are also likely to show exclusion responding when visual arbitrary stimuli A1 and A2 are presented as samples with two novel visual stimuli as comparisons, D1 and D2. Participants received training in A-A, N1-A1, N2-A2, N1-B1 and N1-C1 matching. When participants demonstrated (1) N2-B2 and N2-C2 matching via exclusion, and (2) matching performances among A, B, and C during testing, which would indicate the formation of two ABCN stimulus classes, they received A1-D1 training and subsequent A2-D2 testing. If participants maintained A1-D1 and showed A2-D2 during testing, apparently excluding D1 with sample A2, they were expected to show emergent matching performances between A, B, and C stimuli and D1 and D2 (e.g., D-B, C-D, N-D): this would provide evidence for the formation of stimulus equivalence classes (Study question 4).

8. Method

Two groups of four children participated. Group 1 (Participants 1–4) consisted of two girls and two boys between 61 and 62 months (mean age: 61.5 months). Group 2 (Participants 5–8) consisted of two girls and two boys between 58 and 63 months (mean age: 60.75 months). Setting and sessions were the same as in Experiment 1.

Reliability scores were obtained for about 55 % of the sessions. Instructions and reinforcement procedures were the same as in Experiment 1. Stimuli that were used in Experiment 3 are shown in Fig. 3. Other materials were the same as in the previous experiments.

Fig. 1 (bottom row) illustrates the trained (solid arrows) and tested (broken arrows) relations. First, all children were successively taught N-A (N1-A1, N2-A2), and N1-B1 matching. The N-A task was trained in the same way as in Experiment 2; N1-B1 was trained in the same way as N1-C1 matching in Experiment 2, with the same mastery criteria applied. When participants passed the mixed baseline training, they received A-A identity-matching training in the same way as in the previous experiments. Thereafter, participants received maintenance testing of the trained relations and N2-B2 probes; sessions consisted of two blocks of eight test trials each (two A-A, two N-A, two N1-B1, and two N2-B2 probes) and two blocks of six training trials each (two A-A, two N-A, and two N1-B1 trials). When participants maintained baseline performances and showed N2-B2 matching on at least 3/4 probes, they received A-B and B-A probes that tested for emergent arbitrary-matching performances between A and B stimuli on basis of shared relations with N1 or N2. Emergent B-A and A-B matching was tested in the same way as in Experiment 2.

When participants passed the A-B and B-A probes, Group 1 participants received N1-C1 training; Group 2 participants received N1-C2 training (stimulus counterbalancing). N1-C1 was trained in the same way as in Experiment 2, with the same mastery criterion applied. After participants mastered N1-C1 training, N2-C2 matching was assessed in the same way as N2-B2 matching with the same mastery criterion. When participants demonstrated N2-C2 matching, they received eight B-C (Block 1) and eight C-B probes (Block 3). Blocks 1 and 2 consisted of six training trials each (two A-A, two N-A, one N1-B1, and one N1-C1 trial); mastery criterion required class-consistent responding on at least 7/8 probe trials and correct responding on at least 11/12 training trials. In the following test session, participants received eight A-C (Block 1) and eight C-A probes (Block 3); Blocks 2 and 4 were the same as during B-C and C-B testing, with the same performance criteria applied. As a consequence of stimulus counterbalancing, for Group 1 participants, responses that were consistent with classes A1B1C1N1 and A2B2C2N2 were considered accurate, for Group 2 participants, responses that were consistent with classes A1B1C2N1 and A2B2C1N2 were considered accurate.

When children showed arbitrary-matching performances between A,

B, and C stimuli, Group 1 participants further received A1-D1 training and Group 2 participants received A1-D2 training. Participants received blocks of 12 training trials. During A1-D1 training, A1 was the sample stimulus and novel stimuli D1 and D2 were comparison stimuli (Fig. 3). On each trial, the experimenter pointed to the sample stimulus and asked the child: "Which one goes with this one?" When a participant selected the designated correct comparison, the experimenter said: "Yes! Take a bead." If the participants selected the designated incorrect comparison, the experimenter said: "No! No bead." Mastery criterion required 11/12 correct responses within one block (maximum three blocks). Thus, participants did not receive A2-D2 (Group 1) or A2-D1 (Group 2) training.

After reliably selecting D1 in the presence of A1, participants received several test sessions. Test sessions consisted of two blocks (Blocks 1 and 3) of 12 test trials each and two blocks (Blocks 2 and 4) of six training trials each (one trial of A-A (A1-A1 in Block 2, A2-A2 in Block 4), N1-A1, N2-A2, N1-B1, N1-C1, and A1-D1 matching). In Block 1, the probe type of interest determined test trial combinations; for example, if the probe of interest involved B-D matching, four B-D probes were mixed with four B-A and four A-D probes. This was done to ascertain criterion responding on probes that were considered prerequisite for emergent matching on the probe trials of interest ("if B then A" and "if A then D", then, "if B then D"). In Block 2, eight probe trials, D-B probes for example, were mixed with four previously examined probes, B-D for example. Mastery criterion on test sessions required accurate responding on 11/12 test trials in each block and on 11/12 training trials.

First, participants received a mix of six A1-D1 and six A2-D2 probes (Block 1). In Block 3, they received eight D-A symmetry probes mixed with four A-D probes. When participants demonstrated the A-D and D-A conditional discriminations, they received four B-D probes mixed with four B-A and four A-D probes (Block 1), followed by eight D-B probes mixed with four B-D probes. Next, participants received four C-D probes mixed with four C-A and four A-D probes (Block 1), followed by eight D-C probes mixed with four C-D probes (Block 3). Finally, the children received two blocks of test trials each consisting of six N-D probes mixed with four A-D and two N-A probes. Group 1 participants were expected to relate D1 to members of class A1B1C1N1 and D2 to members of class A2B2C2N2; Group 2 participants were expected to relate D1 to members of class A2B2C1N2 and D2 to members of class A1B1C2N1.

9. Results and discussion

The experiment consisted of 3728 trials, 2100 training and 1628 test trials. Reliability scores were made on 2112 trials (56.7 %), on 1124 training trials (53.5 %) and on 988 test trials (60.7 %). The experimenter's recordings agreed with the observer's recordings on all trials.

Participants took two blocks of 12 training trials to learn the N-A task. When the N1-B1 relation was trained, all eight participants responded correctly on at least 11/12 trials in the first block. During N-A and N1-B1 mixed training, all participants except Participant 5 reached criterion in one session. In the first session, Participant 5 responded correctly on 19/24 trials (errors on N-A trials only). When this session was repeated, Participant 5 reached criterion. When the A-A task was trained, all participants took one block of 12 trials to reach criterion.

Table 4 shows the results on N2-B2 exclusion probes. During combined maintenance and N2-B2 testing, all participants reached criterion on both test and training trials. On all subsequent test sessions, all participants except Participant 5 maintained criterion responding on baseline training and test trials. Table 4 also shows the results on A-B and B-A probes. In the first session with B-A probes, Participants 1, 2, 3, and 5 demonstrated emergent conditional relations between B and A stimuli on at least 7/8 B-A trials. Participants 4, 6, 7, and 8 showed class-consistent responses on 4/8, 3/8, 6/8, and 5/8 B-A trials, respectively.

When A-B probes were given, all participants except for Participants 2 and 8 reliably showed the expected conditional relations. Participants

Table 4

Experiment 3: Individual results on arbitrary-matching probes for ABCN class formation of sessions in which baseline responding was maintained. Results are presented as number of trials on which the designated comparison was selected/total number of test trials.

Participant no.		1	2	3	4	5	6	7	8
Probes									
N2-B2	3/4	4/4	4/4	4/4	3/4	4/4	4/4	4/4	3/4
	4		4		4				
B-A	8/8	8/8	7/8	4/8; 8/8	8/8	3/8; 8/8	6/8; 8/8	5/8; 8/8	5/8; 8/8
	8	8	8	8	8	8	8	8	8
A-B	8/8	5/8; 8/8	8/8	7/8	8/8	8/8	7/8	5/8; 7/8	5/8; 7/8
	8	8	8	8	8	8	8	8	8
N2-C2	4/4	4/4	4/4	3/4	3/4	4/4	3/4	4/4	4/4
	4		4		4				
B-C	8/8	8/8	8/8	7/8	8/8	7/8; 8/8	7/8	7/8	7/8
	8	8	8	8	8	8	8	8	8
C-B	8/8	8/8	7/8	8/8	6/8	6/8; 8/8	8/8	7/8	7/8
	8	8	8	8	8	8	8	8	8
A-C	8/8	8/8	7/8	8/8	-	8/8	8/8	7/8	7/8
	8	8	8						
C-A	7/8	8/8	8/8	8/8	-	8/8	7/8	8/8	8/8
	8	8	8						

Note. "-" = The participant did not receive these probes.

2 and 8 responded correctly on 5/8 A-B probes while they responded class-consistently on mixed B-A probes. Following A-B testing, a second B-A test session was conducted for Participants 4, 6, 7, and 8, in which responding was class-consistent on 8/8 B-A probes. Next, a second A-B test session was conducted for Participants 2 and 8, in which Participant 2 responded accurately on 8/8 and Participant 8 on 7/8 A-B probes. Thus, all participants showed emergent arbitrary-matching performances among A and B stimuli. During N1-C1 training, all participants reached criterion within one block of 12 trials. During combined maintenance and N2-C2 testing, all participants reached criterion on both test and training trials. On B-C and C-B probes, all participants except Participant 6 showed the expected matching performances immediately in a reliable way. In the first B-C and C-B test session, Participant 6 responded correctly on 7/8 B-C probes and on 6/8 C-B probes. When this test session was repeated, Participant 6 responded perfectly on all test and training trials. When participants received A-C and C-A probes, all participants except Participant 5 showed the expected arbitrary-matching performances. For unclear reasons, Participant 5 did not maintain criterion responding on baseline test and training trials during two successive AC/CA test sessions. Therefore, Participant 5 was excluded from further participation.

The remaining seven participants took one to two blocks of A1-D1

Table 5

Experiment 3. Individual results on exclusion and arbitrary matching probes involving D stimuli. The results are presented as the number of trials on which the designated comparison was selected/total amount of test trials.

Participant no.		1	2	3	4	5	6	7	8
Test/Probes									
A1-D1	6/6	5/6	6/6	5/6; 6/6	6	-	6/6	6/6	5/6
A2-D2	6/6	6/6	6/6	3/6; 5/6	6	-	6/6	5/6	6/6
D-A	8/8	8/8	8/8	7/8	-	6/8	8/8	7/8	7/8
B-D	8/8	7/8	8/8	8/8	-	8/8	7/8	8/8	8/8
D-B	8/8	7/8	8/8	8/8	-	8/8	7/8	7/8	7/8
C-D	8/8	8/8	8/8	8/8	-	8/8	8/8	8/8	8/8
D-C	8/8	6/8	8/8	7/8	-	8/8	8/8	8/8	8/8
N-D	12/12	12/12	12/12	11/12	-	12/12	11/12	11/12	11/12
	12	12	12			12	12	12	12

Note. "-" = the participant did not receive these probes.

training to reach criterion. Table 5 shows the results on probes involving D1 and D2. When they received A1-D1 and A2-D2 test trials (Block 1), all participants except Participant 4 matched A-D at criterion level. Participant 4 reached criterion when he received the test trials of Block 1 a second time. On D-A probes, all participants except Participant 6 showed the expected matching performances on at least 7/8 trials of each probe type, while maintaining A1-D1 and A2-D2 matching. Participant 6 responded accurately on 6/8 D-A probes (75 %). On B-D and D-B probes, all participants reliably showed the expected matching performances while maintaining matching on other probes and criterion responding on training trials. On C-D and D-C probes, all participants except Participant 2 showed expected matching performances on at least 7/8 probe trials. Participant 2 responded class-consistently on 8/8 C-D probes and 6/8 D-C probes (75 %). Finally, when N-D probes (N1-D1 and N2-D2) were presented, all participants reliably selected the accurate D comparison on basis of the experimenter's question "Which one is correct?" (N1) and "Which one is incorrect?" (N2).

The results of Experiment 3 can be summarised as follows: first, after being taught N1-A1, N2-A2, N1-B1, and N1-C1, all participants showed N2-B2 and N2-C2 matching via exclusion and seven participants subsequently showed emergent matching performances among A, B, and C stimuli. This finding suggests that in Experiment 2, N2-B2 training may have been redundant.

Second, when the remaining seven participants received training of A1-D1 matching, they all demonstrated A2-D2 by exclusion testing and subsequently showed emergent matching performances among D and A, B, and C stimuli, respectively. Although on the first probes (B-A probes), the expected arbitrary-matching performances emerged more gradually compared with the results in Experiment 2, participants did show these performances within two sessions. The gradual establishment of B-A and A-B matching performances might also result from the fact that (1) a relatively small number of N2-B2 test trials were conducted or (2) fewer matching tasks were trained compared to Experiment 2. Matching performances involving C1 and C2, however, readily emerged despite the small number of N2-C2 probes. The results suggest that not only the spoken words N1 and N2 were sufficiently distinct to generate, for example, N2-B2 matching after N1-B1 training, but also the visual stimuli such as A1 and A2 can generate exclusion responding. It should be noted, however, that A1 and A2 were already related to N1 and N2, respectively, which might have contributed to the distinctiveness of A1 versus A2. To what extent an analysis of stimulus variables is necessary to predict children's performances on exclusion probes remains to be investigated. In conclusion, the results on all arbitrary-matching probes show that seven participants were able to conditionally relate not only visual A, B, C, and D stimuli, but also the novel D stimuli upon N1 and N2. These emergent matching performances suggest the formation of two five-member stimulus equivalence classes.

10. General discussion

The current study reported the results of a systematic series of experiments regarding the degree to which exclusion responding testing or training conditions in will predict the emergence of equivalence relations in preschool children. Results show that relations established under exclusion conditions may further generate emergent matching performances in young children that are indicative of equivalence class formation. In Experiment 1, only N1-A1 and N1-B1 relations were taught explicitly. Subsequently, even though participating children reliably demonstrated N2-A2 and the N2-B2 relations through exclusion under testing conditions, corresponding A2-B2 and B2-A2 matching performances did not reliably emerge. On exclusion probes in all three experiments, participants consistently selected the novel comparison with N2 as sample, suggesting conditional control by N stimuli. Furthermore, in Experiments 2 and 3 the relations that were established under exclusion conditions brought about other emergent matching performances. The results of Experiment 1, however, showed that

children were less likely to demonstrate emergent conditional matching performances other than those demonstrated on exclusion probes.

In all experiments except Experiment 1, the results of B-A and A-B arbitrary-matching probes were indicative of the formation of two ABN equivalence classes. In Experiments 2 and 3, the established stimulus classes were extended with novel stimuli that participated in relations with class members via exclusion responding. In other words, if one of two comparisons is selected in the presence of a defined sample during training and the other, undefined, comparison is selected in the presence of another defined sample during testing, then both comparison stimuli may come to participate in a derived stimulus relation with their respective samples. This finding can be considered an addition to previous findings and recommendations on how to achieve stable, sample- S^+ relations following responding by exclusion (e.g., Carr, 2003; Costa et al., 2001; Domeniconi et al., 2007; Ferrari et al., 1993, 2008; Langsdorff et al., 2017; McIlvane and Stoddard, 1981; McIlvane et al., 1992a, 1992b, 2009; Wilkinson and McIlvane, 1997).

The results from Experiment 1 clearly show that matching performances that are established under exclusion test conditions do not necessarily or automatically generate emergent matching performances that reflect the formation of arbitrary stimulus classes. This finding is in line with previous research findings with preschool participants (e.g., Smeets and Barnes, 1997). In addition to previous recommendations regarding methodological arrangements to encourage stable stimulus- S^+ performances (see Langsdorff et al. (2017) for an overview), the current results suggest a number of issues should be taken into account. The first issue relates to whether the sample stimuli that are presented on exclusion probes already participated in previously trained relations (i.e., are samples already defined by reinforcement contingencies) or not. The current results suggest that, with preschool children, only when the sample stimuli are defined by reinforcement contingencies are participants likely to show the expected matching performances. The second issue relates to the number of conditional discriminations that are trained preceding the critical B-A and A-B testing phases. The current results suggest that the more relations trained, the more quickly (as opposed to gradually or not at all) participants demonstrate expected arbitrary matching performances. This issue follows from comparing the results for baseline training of the three experiments and the pilot study. After all, the number of trained matching tasks preceding B-A and A-B testing differed between the several experiments: (1) four matching tasks in the pilot study and in Experiment 2 (result: participants immediately showed the expected arbitrary matching performances); (2) three matching tasks in Experiment 3 (result: participants showed the emergent performances rather gradually); and (3) two matching tasks in Experiment 1 (result: only three participants showed the expected matching performances). This observation (i.e., the more conditional relations trained, the quicker expected arbitrary matching performances can be observed) implies that extended baseline training, possibly even with novel, non-experimental stimuli, should increase the number of participants showing the expected matching performances. This conclusion is supported by the results in a study with children with autism by Carr (2003) who showed that reinforced exclusion trials with four novel word-item stimulus pairs facilitated nonreinforced exclusion responding. The participants in Carr's study, who had initially failed tests for exclusion, were exposed to reinforced exemplars of exclusion responding and nonreinforced exclusion trials with unknown items within the same block. It remains to be seen, therefore, whether emergent exclusion performances necessarily depend on reinforced and nonreinforced trials being presented within the same trial block (e.g., Wilkinson and Green, 1998). If exclusion responding is facilitated through multiple-exemplar training (e.g. Luciano et al., 2007) then these findings, combined with those of the present study, suggest that "exclusion performances result when participants learn generalised rejection of any comparison that is not in the same experimentally defined stimulus class as the sample" (McIlvane et al., 1988, p.492). Determining the number of reinforced or nonreinforced exemplars that

are necessary before generalised exclusion responding is demonstrated is also an interesting and important empirical question requiring further investigation (see also Barnes-Holmes et al., 2001; Luciano et al., 2007). Finally, the amount of participant exposure to (reinforced) examples of S^+ and S^- control may also play a role or even provide an equally valid explanation for the finding that ABN classes were not established in Experiment 1, but were established in Experiments 2 and 3. This suggestion is not only in line with the study by da Costa et al. (2013), but also with the findings in a recent study by Plazas and Villamil (2018). In three experiments involving first-semester students, they studied the formation of equivalence relations among stimuli that that previously had been related by exclusion only. Plazas and Villamil found that, in line with previous studies, responding by exclusion did not automatically generate stimulus equivalence relations for all participant (only 66 % of adolescent/adult participants showed the expected stimulus classes over the three experiments). In Experiment 3 of that study, however, they found that increased exposure to exclusion trials might have strengthened conditional relations between stimuli that were related by exclusion. Within the context of conditional discrimination with regard to the current study, it may be the case that participants were insufficiently exposed (in terms of number of trials) to reinforced examples of S^- control and/or S^+ control, defined by sample-comparison reinforcement contingencies. Compared to the study by Plazas and Villamil, where equivalence classes were formed exclusively as a result from negative control (thus in the absence of reinforcement), in the current experiments, responding by exclusion was followed by consequences on a limited number of training trials before equivalence classes were formed; thereafter, classes could be expanded with new stimuli that were selected on exclusion probes. In the current experiments, our young participants may have required richer schedules of reinforcement and/or may have been more sensitive to changes in contingencies (e.g. from training to testing conditions) compared to adult participants, which may offer an possible explanation of why our participants still needed some training (with consequences) of responding by exclusion. Since this was not further explored in the study, this limitation would be highly interesting to further address in future studies. Although the results of both Experiments 2 and 3 clearly demonstrated that manipulations of reinforcement contingencies for N2 generated more stable conditional discriminations on all probes, it cannot be excluded that, compared to Experiment 1, in Experiments 2 and 3 the increased amount of reinforced exposure to not only S^+ but particularly S^- control might have been sufficient for deriving all associated emergent relations. The latter may also apply as an possible explanation for why the training of N2-B2 in addition to N2-A2 relations in Experiment 2 was shown to be redundant in Experiment 3; i.e., participants in Experiment 3 may have had sufficient reinforced exposure both to S^+ and S^- control elements that were necessary for deriving the associated emergent relations in the context of conditional discriminations. Future experimentation might further explore this issue including how may possibly affect equivalence properties of the derived relations (Perez et al. (2017).

In Experiment 4, the results of the N-D probes suggest that the D stimuli acquired the same functions as their respective class members through ABCD class formation. It remains to be seen whether stimuli that have acquired the same function but do not participate in any stimulus class, become conditionally related via shared relations with that function.

The use of spoken words "correct" (N1) and "incorrect" (N2) in the present study may be rather unusual compared to other studies on conditional discrimination learning because selections of "correct" as well as "incorrect" stimuli could produce reinforcement. In two-choice simple discrimination tasks, for example, "correct" stimuli usually predict reinforcement but "incorrect" stimuli do not (e.g., de Rose et al., 1988a, b; Smeets, 1991). In Experiment 1, A1 and B1 were, theoretically, potential members of one class, and A2 and B2 were potential members of another class. After all, A1 and B1 were both defined stimuli, as a result of training, and A2 and B2 were both undefined. In

the study by [Plazas and Villamil \(2018\)](#), participants were faced with a similar challenge to relate undefined stimuli that were selected by exclusion. A pilot study described by [McIlvane et al. \(1992a\)](#) also addressed this possibility. In that study, participants with intellectual difficulties received training involving two simple discriminations tasks, A1 (S^+) versus A2 (S^-), and B1 (S^+) versus B2 (S^-). On subsequent A-B and B-A arbitrary-matching probes, participants were expected to relate A1-B1 and A2-B2. The results showed, however, that the participants usually selected the stimuli that functioned as S^+ , regardless of which sample was presented, so-called restrictive stimulus control (e.g. [Bickel et al., 1984](#); [Dube, 1997](#); [Dube and McIlvane, 1999](#)). Similar response patterns were on MTS probes, i.e. selecting previous S^+ stimuli only, were reported for five participants in Experiment 1 of the present study (despite demonstrating matching N1-A1, N1-B1, N2-A2, and N2-B2, the prerequisite conditional relations for subsequent B-A and A-B matching performances.) Experiments 2 and 3, however, showed one possible pathway of overcoming this so-called 'restrictive stimulus control' by training participants to master at least three conditional relations training (e.g. N1-A1, N2-A2, and N1-B1) before introducing exclusion trials and further relation training with the aim of rapidly expanding the two arbitrary stimulus classes (A1B1N1 and A2N2). The results of the current study also have implications for further psycho-linguistic experimentation and future educational practices.

First, the current findings have relevance to future experimental psycholinguistic studies since it allows for the experimental study (and possible explanation) of children's rapid increase of vocabulary expansion during the second year of their life (e.g. [Ganger and Brent, 2004](#)). The study by [Domeniconi et al. \(2007\)](#) already showed that participants as young as 2–3 years old, are capable of exclusion responding, one of the reasons that exclusion responding can be effectively used in teaching programmes for young learner's reading and spelling behaviours, such as picture naming, echoing, copying, and writing behaviours as well as matching to sample of printed and dictated words, and pictures (e.g., [Dixon, 1977](#); [Domeniconi et al., 2007](#); [de Souza et al., 2009](#)). Future studies might explore whether vocabulary spurt that is often observed in these young learners, may (partly) be the result of emergent stimulus-stimulus relations, or could even be retrospectively explained by further studying the minimal age at which children start using responding by exclusion. Emergent stimulus-stimulus relations following exclusion responding may also have potential benefits daily life teaching situations, in class or at home.

Teaching may be more efficient when only those conditional relations are trained or taught that are necessary for arbitrary stimulus classes to emerge and subsequent use of responding by exclusion will lead to relatively more and increasing number of emergent relations (leading to larger stimulus equivalence classes). To illustrate: in Experiment 3, two ABCDN stimulus classes were established with participants showing 16 arbitrary conditional relations between stimuli in each class (32 in total). Of these 32 conditional relations, 27 were emergent, only 5/32 relations were taught explicitly, instead of the typical 8/32. The number of conditional relations is normally even higher, but due to the nature of the N stimuli, some conditional relations could not be assessed, e.g., A-N, B-N etc. It is hypothesised, however, that had it been possible to assess all conditional relations, then children could have demonstrated up to 40 conditional relations (2×20 for each class) of which only 5/40 were trained, thereby reducing the number of trained relations from 20 % (8/40) to 12.5 % (5/40). When the goal is to make teaching more efficient, which is to minimise the number of trained and to maximize the number of emergent relations, e.g. when teaching language skills, educators could critically evaluate whether any elements of their teaching is redundant in order to obtain the same intended learning outcomes (i.e. the total number of equivalent relations between verbal and visual stimuli).

Recently, researchers have questioned whether or not the currently reported experiments and similar experiments on equivalence in the field, show a clear demonstration of experimental control; after all,

within each experiment of the current study, contingency manipulations were the same for all participants and none of the experiments consisted of a parallel running control condition, making it difficult to attribute changes in subjects' responses to manipulation of contingencies. In the current study, however, one could argue that changes in responses on test trials, were in fact the result of manipulations of reinforcement contingencies; after all, a series of experiments were conducted (a pilot study, followed by Experiments 1, 2, and 3) in which, predominantly and systematically, only reinforcement contingencies were manipulated, while keeping other factors constant (e.g. age and gender of participants, the use of naïve participants with regard to previous experiment experience, training and testing format, number of trials, settings and sessions, mastery criteria etc.). One could argue that Experiment 1 served as a control condition for the results of the pilot study and that a vertical comparison of designs of all experiments (pilot and Experiments 1, 2, and 3: see overview [Fig. 1](#)) allows for interpretation of the findings in terms of experimental control. In addition, however, it is recommended that future experiments on stimulus equivalence would also consider taking measures to more convincingly demonstrate experimental control within each experiment; for example, by using a (concurrent or nonconcurrent) multiple baseline design, by introducing changes in reinforcement contingencies in one condition and not in the other (concurrent) or at different intervals across participants (nonconcurrent) ([Christ, 2007](#)). Other measures that could be taken to increase the internal validity of equivalence experiments might involve randomisation of stimuli and designated S^+ and S^- across participants. In addition, the use of only two comparisons, rather than three or more, may be seen as a limitation of the current experiments ([Sidman, 1987](#); [Carrigan and Sidman, 1992](#)).

The decision to use only two comparisons was, however, considered to be an appropriate choice, given the use of the two opposite verbal labels "correct" and "incorrect". Our experiments, however, were designed in such a way that they would ensure equal chances of positive and negative control. Future studies using not only different stimuli but also offering more than just two comparisons, however, might yield further knowledge on how responding by exclusion might benefit learning and teaching in applied settings, although using more than two comparisons might also further complicate response interpretation ([Boelens, 2002](#)).

In conclusion, the present study contributes to previous findings on children's exclusion responding in matching-to-sample with regard to the expansion of arbitrary stimulus classes with novel stimuli that are selected in the presence of the sample stimuli on exclusion probes. Thus, performances on exclusion probes may not only provide information on the nature of participants' matching performances, as in previous studies (e.g., [Dixon and Dixon, 1978](#); [Stromer and Osborne, 1982](#)), but may also be used to expand existing arbitrary stimulus classes. The necessary and sufficient conditions required to obtain emergent matching performances based on relations that are established under exclusion conditions, as well as the long-term stability of these relations, should be subject to further study considering their relevance to both psycholinguistic as well as experimental and applied behaviour-analytic research.

Author note

This research was conducted as part of the first author's PhD research at Leiden University, The Netherlands. This study was presented at the 3rd International Conference of the Association for Behaviour Analysis International, November 25–27, 2005, Beijing (China). A summary of this paper also was presented to the Annual Association for Behaviour Analysis Convention, Washington DC (1995), and a two-column summary was published in the *Experimental Analysis of Human Behavior Bulletin*, 13(1), 3, both as part of the ABA Special Interest Group for the Experimental Analysis of Human Behavior 1995 Student Paper Award.

Funding details

The authors received no financial support for the research, authorship, and/or publication of this article.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Active informed consent was obtained from all individual participants (and their parents/legal guardians) included in the study.

Availability of data and materials

Data supporting our findings (incl. Supplementary Materials) can be obtained by contacting the first author.

CRediT authorship contribution statement

Jacqueline J. Schenk: Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Data curation, Visualisation. **Mickey Keenan:** Writing - review & editing, Visualisation. **Harrie H. Boelens:** Writing - review & editing. **Simon Dymond:** Writing - review & editing. **Paul M. Smeets:** Conceptualization, Validation, Supervision, Investigation, Resources, Project administration.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgement

The authors would like to thank all the children who participated in this study.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.beproc.2021.104341>.

References

- Aust, U., Range, F., Steurer, M., Huber, L., 2008. Inferential reasoning by exclusion in pigeons, dogs, and humans. *Anim. Cogn.* 1 (4), 587–597. <https://doi.org/10.1007/s10071-008-0149-0>.
- Barnes, D., Browne, M., Smeets, P.M., Roche, B., 1995. A transfer of functions and a conditional transfer of functions through equivalence relations in three to six year old children. *Psychol. Rec.* 45, 405–430. <https://doi.org/10.1007/BF03395151>.
- Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., Smeets, P.M., 2001. Exemplar training and a derived transformation of function in accordance with symmetry. *Psychol. Rec.* 51, 287–308. <https://doi.org/10.1007/BF03395400>.
- Bates, E., 1979. *The Emergence of Symbols: Cognition and Communication in Infancy*. Academic Press, New York.
- Beran, M.J., Washburn, D.A., 2002. Chimpanzee responding during matching to sample: control by exclusion. *J. Exp. Anal. Behav.* 78 (3), 497–508. <https://doi.org/10.1901/jeab.2002.78-497>.
- Bickel, W.K., Stella, M.E., Etzel, B.C., 1984. A reevaluation of stimulus overselectivity: restricted stimulus control or stimulus control hierarchies. *J. Autism Dev. Disord.* 14, 137–157. <https://doi.org/10.1007/BF02409657>.
- Boelens, H., 2002. Studying stimulus equivalence: defense of the two-choice procedure. *Psychol. Rec.* 52, 305–314. <https://doi.org/10.1007/BF03395432>.
- Carey, S., Bartlett, E., 1978. Acquiring a single new word. *Papers Rep. Child Lang. Dev.* 15, 17–29.
- Carr, D., 2003. Effects of exemplar training in exclusion responding on auditory-visual discrimination tasks with children with autism. *J. Appl. Behav. Anal.* 36 (4), 507–524. <https://doi.org/10.1901/jaba.2003.36-507>.
- Carr, D., Felce, D., 2000. Application of stimulus equivalence to language intervention for individuals with severe linguistic disabilities. *J. Intellect. Dev. Disabil.* 25, 181–205. <https://doi.org/10.1080/13269780050144262>.
- Carrigan, P.F., Sidman, M., 1992. Conditional discrimination and equivalence relations: a theoretical analysis of control by negative stimuli. *J. Exp. Anal. Behav.* 8 (1), 183–204. <https://doi.org/10.1901/jeab.1992.58-183>.
- Christ, T.J., 2007. Experimental control and threats to internal validity of concurrent and nonconcurrent multiple baseline designs. *Psychol. Sch.* 44, 451–459. <https://doi.org/10.1002/pits.20237>.
- Cippola, N.S., Domeniconi, C., Machado, A., 2014. Responding by exclusion in temporal discrimination tasks. *J. Exp. Anal. Behav.* 101 (2), 215–229. <https://doi.org/10.1002/jeab.71>.
- Costa, A.R.A., Wilkinson, K.M., McIlvane, W.J., de Souza, D.G., 2001. Emergent word-object mapping by children: further studies using the blank comparison technique. *Psychol. Rec.* 51 (3), 343–355. <https://doi.org/10.1007/BF03395403>.
- Costa, A.R.A., Grisante, P., Domeniconi, C., de Rose, J.C., de Souza, D.G., 2013. Naming new stimuli after selection by exclusion. *Revista Paidéia* 55, 217–224. <https://doi.org/10.1590/1982-43272355201309>.
- Cumming, W.W., Berryman, R., 1965. The complex discriminated operant: studies of matching-to-sample and related problems. In: Mostofsky, D.I. (Ed.), *Stimulus Generalization*. Stanford University Press, Stanford, CA, pp. 284–330.
- de Rose, J.C., McIlvane, W.J., Dube, W.V., Galpin, V.C., Stoddard, L.T., 1988a. Emergent simple discrimination established by indirect relation to differential consequences. *J. Exp. Anal. Behav.* 50, 1–20. <https://doi.org/10.1901/jeab.1988.50-1>.
- de Rose, J.C., McIlvane, W.J., Dube, W.V., Stoddard, L.T., 1988b. Stimulus class formation and functional equivalence in moderately retarded individuals' conditional discrimination. *Behav. Processes* 17, 167–175. [https://doi.org/10.1016/0376-6357\(88\)90033-2](https://doi.org/10.1016/0376-6357(88)90033-2).
- de Souza, M.F., Schmidt, A., 2014. Responding by exclusion in Wistar rats in a simultaneous visual discrimination task. *J. Exp. Anal. Behav.* 102 (3), 346–352. <https://doi.org/10.1002/jeab.106>.
- de Souza, D.G., de Rose, J.C., Domeniconi, C., 2009. Applying relational operants to reading and spelling. In: Rehfeldt, R.A., Barnes-Holmes, Y. (Eds.), *Derived Relational Responding: Applications for Learners With Autism and Other Developmental Disabilities*. New Harbinger, Oakland, CA, pp. 173–207.
- Dixon, L.S., 1977. The nature of control by spoken words over visual stimulus selection. *J. Exp. Anal. Behav.* 27, 433–442. <https://doi.org/10.1901/jeab.1977.27-433>.
- Dixon, M.H., Dixon, L.S., 1978. The nature of standard control in children's matching to sample. *J. Exp. Anal. Behav.* 30, 205–212. <https://doi.org/10.1901/jeab.1978.30-205>.
- Dixon, M.H., Dixon, L.S., Spradlin, J.E., 1983. Analysis of individual differences of stimulus control among developmentally disabled children. In: Gadow, K.D., Bailer, I. (Eds.), *Advances in Learning and Behavioral Disabilities*, Vol. 2. JAI Press, New York, pp. 85–110.
- Domeniconi, C., Costa, A.R.A., de Souza, D.G., 2007. Exclusion responding by 2 to 3 year-old children in a play setting. *Psicol. Reflex. Crit.* 20 (2), 342–350.
- Dube, W.V., 1997. Restricted stimulus control and stimulus-reinforcer relations. *Exp. Anal. Hum. Behav. Bull.* 15, 8–11.
- Dube, W.V., McIlvane, W.J., 1999. Reduction of stimulus overselectivity with nonverbal differential observing responses. *J. Appl. Behav. Anal.* 32, 25–33. <https://doi.org/10.1901/jaba.1999.32-25>.
- Dymond, S., Rehfeldt, R., Schenk, J.J., 2005. Nonautomated procedures in derived stimulus relations research: a methodological note. *Psychol. Rec.* 55, 461–481. <https://doi.org/10.1007/BF03395521>.
- Ferrari, C., de Rose, J.C., McIlvane, W.J., 1993. Exclusion vs. selection training of auditory-visual conditional relations. *J. Exp. Child Psychol.* 56, 49–63. <https://doi.org/10.1006/jecp.1993.1025>.
- Ferrari, C., Rose, J.C., McIlvane, W.J., 2008. A comparison of exclusion and trial-and-error procedures: primary and secondary effects. *Exp. Anal. Hum. Behav. Bull.* 29, 9–16.
- Ganger, J., Brent, M.R., 2004. Reexamining the vocabulary spurt. *Dev. Psychol.* 40 (4), 621–632. <https://doi.org/10.1037/0012-1649.40.4.621>.
- Grassmann, S., Stracke, M., Tomasello, M., 2009. Two-year-olds exclude novel objects as potential referents of novel words based on pragmatics. *Cognition* 112 (3), 488–493. <https://doi.org/10.1016/j.cognition.2009.06.010>.
- Grassmann, S., Schulze, C., Tomasello, M., 2015. Children's level of word knowledge predicts their exclusion of familiar objects as referents of novel words. *Front. Psychol.* 6, 1200. <https://doi.org/10.3389/fpsyg.2015.01200>.
- Johnson, C., Sidman, M., 1993. Conditional discrimination and equivalence relations: control by negative stimuli. *J. Exp. Anal. Behav.* 59 (2), 333–347. <https://doi.org/10.1901/jeab.1993.59-333>.
- Kato, O.M., de Rose, J.C., Faleiros, P.B., 2008. Topography of responses in conditional discrimination influences formation of equivalence classes. *Psychol. Rec.* 58, 245–267. <https://doi.org/10.1007/BF03395614>.
- Langsdorff, L.C., Domeniconi, C., Schmidt, A., Gomes, C.G., de Souza, D., 2017. Learning by exclusion in individuals with autism and Down syndrome. *Psicol. Reflexão e Crítica* 30, 9. <https://doi.org/10.1186/s41155-017-0064-x>.
- Luciano, C., Becerra, I.G., Valverde, M.R., 2007. The role of multiple-exemplar training and naming in establishing derived equivalence in an infant. *J. Exp. Anal. Behav.* 87, 349–365. <https://doi.org/10.1901/jeab.2007.08-06>.
- McIlvane, W.J., Stoddard, T., 1981. Acquisition of matching to sample performances in severe retardation: learning by exclusion. *J. Ment. Defic. Res.* 25 (1), 33–48. <https://doi.org/10.1111/j.1365-2788.1981.tb00901.x>.
- McIlvane, W.J., Bass, R.W., O'Brien, J.M., Genovay, B.J., Stoddard, L.T., 1984. Spoken and signed naming of foods after receptive exclusion training in severe retardation.

- Appl. Res. Ment. Retard. 5 (1), 1–27. [https://doi.org/10.1016/S0270-3092\(84\)80016-8](https://doi.org/10.1016/S0270-3092(84)80016-8).
- McIlvane, W.J., Kledaras, J.B., Munson, L.C., King, K.A., de Rose, J.C., Stoddard, L.T., 1987. Controlling relations in conditional discrimination and matching by exclusion. *J. Exp. Anal. Behav.* 48, 187–208. <https://doi.org/10.1901/jeab.1987.48-187>.
- McIlvane, W.J., Munson, L.C., Stoddard, L.T., 1988. Some observations on control by spoken words in children's conditional discrimination and matching by exclusion. *J. Exp. Child Psychol.* 45, 472–495. [https://doi.org/10.1016/0022-0965\(88\)90043-4](https://doi.org/10.1016/0022-0965(88)90043-4).
- McIlvane, W.J., Dube, W.V., Kledaras, J.B., de Rose, J.C., Stoddard, L.T., 1992a. Stimulus reinforcer relations and conditional discrimination. In: Hayes, S.C., Hayes, L.J. (Eds.), *Understanding Verbal Relations*. Context Press, Reno, NV, pp. 43–67.
- McIlvane, W.J., Kledaras, J.B., Lowry, M.J., Stoddard, L.T., 1992b. Studies of exclusion in individuals with severe mental retardation. *Res. Dev. Disabil.* 13 (6), 509–532. [https://doi.org/10.1016/0891-4222\(92\)90047-A](https://doi.org/10.1016/0891-4222(92)90047-A).
- Meehan, E.F., 1995. Emergence by exclusion. *Psychol. Rec.* 45, 133–165.
- Mervis, C.B., Bertrand, J., 1994. Acquisition of the novel-name-nameless-category (N3C) principle. *Child Dev.* 63, 1646–1662. <https://doi.org/10.2307/1131285>.
- Nazzi, T., Bertoncini, J., 2003. Before and after the vocabulary spurt: two modes of word acquisition? *Dev. Sci.* 6 (2), 136–142. <https://doi.org/10.1111/1467-7687.00263>.
- Perez, W.F., Tomanari, G.Y., Vaidya, M., 2017. Effects of heterogeneous controlling relations on tests of transitivity and equivalence: an exploratory study. *Eur. J. Behav. Anal.* 18 (2), 251–263. <https://doi.org/10.1080/15021149.2017.1309957>.
- Plazas, E.A., Villamil, C., 2018. Formation of new stimulus equivalence classes by exclusion. *J. Exp. Anal. Behav.* 109 (2), 380–393. <https://doi.org/10.1002/jeab.322>.
- Schenk, J.J., 1995. Complex stimuli in non-reinforced simple discrimination tasks: emergent simple and conditional discriminations. *Psychol. Rec.* 45, 477–494. <https://doi.org/10.1007/BF03395156>.
- Sidman, M., 1987. Two choices are not enough. *Behav. Anal.* 22, 11–18.
- Sidman, M., 1992. Equivalence relations: some basic considerations. In: Hayes, S.C., Hayes, L.J. (Eds.), *Understanding Verbal Relations*. Context Press, Reno, NV, pp. 15–27.
- Sidman, M., 1994. *Equivalence Relations and Behavior: A Research Story*. Authors Cooperative, Boston, MA.
- Sidman, M., 2000. Equivalence relations and the reinforcement contingency. *J. Exp. Anal. Behav.* 74, 127–146. <https://doi.org/10.1901/jeab.2000.74-127>.
- Sidman, M., Tailby, W., 1982. Conditional discrimination vs. matching to sample: an expansion of the testing paradigm. *J. Exp. Anal. Behav.* 37, 5–22. <https://doi.org/10.1901/jeab.1982.37-5>.
- Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., Carrigan, P., 1982. A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. *J. Exp. Anal. Behav.* 37, 23–44. <https://doi.org/10.1901/jeab.1982.37-23>.
- Smeets, P.M., 1991. Emergent simple discrimination in children: Transfer of stimulus control under nonreinforced conditions. *Quarterly Journal of Experimental Psychology* 43B, 361–388.
- Smeets, P.M., Barnes, D., 1997. Emergent conditional discriminations in children and adults: stimulus equivalence derived from simple discriminations. *J. Exp. Child Psychol.* 66 (1), 64–84. <https://doi.org/10.1006/jecp.1997.2377>.
- Smeets, P.M., Barnes, D., Schenk, J.J., Darcheville, J.C., 1996. Emergent simple discriminations and conditional relations in children, adults with mental retardation, and normal adults. *Q. J. Exp. Psychol.* 49B, 201–219. <https://doi.org/10.1080/713932626>.
- Stromer, R., 1986. Control by exclusion in arbitrary matching-to-sample. *Anal. Interv. Dev. Disabil.* 6, 59–72. [https://doi.org/10.1016/02704684\(86\)90006-6](https://doi.org/10.1016/02704684(86)90006-6).
- Stromer, R., 1989. Symmetry of control by exclusion in humans' arbitrary matching-to-sample. *Psychol. Rep.* 64, 915–922. <https://doi.org/10.2466/pr0.1989.64.3.915>.
- Stromer, R., Osborne, J.G., 1982. Control of adolescents' arbitrary matching-to sample by positive and negative stimulus relations. *J. Exp. Anal. Behav.* 37, 329–348. <https://doi.org/10.1901/jeab.1982.37-329>.
- Tomonaga, M., 1993. Tests for control by exclusion and negative stimulus relations of arbitrary matching to sample in a "symmetry-emergent" chimpanzee. *J. Exp. Anal. Behav.* 59, 215–229. <https://doi.org/10.1901/jeab.1993.59.215>.
- Wilkinson, K.M., Green, G., 1998. Implications of fast mapping for vocabulary expansion in individuals with mental retardation. *Augment. Altern. Commun.* 14, 162–170.
- Wilkinson, K.M., McIlvane, W.J., 1997. Blank comparison analysis of emergent symbolic mapping by young children. *J. Exp. Child Psychol.* 67 (2), 115–130. PMID: 9388802.
- Wilkinson, K.M., Dube, W.V., McIlvane, W.J., 1996. A cross-disciplinary perspective on studies of rapid word mapping in psycholinguistics and behaviour analysis. *Dev. Rev.* 16, 125–148.
- Wilkinson, K.M., de Souza, D.G., McIlvane, W.J., 2000. As origens da exclusão. *Temas Em Psicol.* 8 (2), 195–203. Retrieved from. http://pepsic.bvsalud.org/scielo.php?pid=S1413-389X2000000200009&script=sci_arttext.
- Wilkinson, K.M., Rosenquist, C., McIlvane, W.J., 2009. Exclusion learning and emergent symbolic category formation in individuals with severe language impairments and intellectual disabilities. *Psychol. Rec.* 59 (2), 187–206. PMID: PMC2818096.
- Zaine, I., Domeniconi, C., de Rose, J.C., 2016. Exclusion performance and learning by exclusion in dogs. *J. Exp. Anal. Behav.* 105 (3), 362–374. <https://doi.org/10.1002/jeab.209>.