The Psychological Record, 2003, 53, 349-371

EQUIVALENCE-EQUIVALENCE AS A MODEL OF ANALOGY: FURTHER ANALYSES

FRANCK CARPENTIER and PAUL M. SMEETS
Leiden University

DERMOT BARNES-HOLMES
National University of Ireland at Maynooth

Equivalence-equivalence is assumed when training of A-B and A-C matching tasks not only leads to matching same-class B and C stimuli but also to matching BC compounds with same-class elements (e.g., B1C1-B2C2) and with different-class elements (e.g., B1C2-B2C3). Like classical analogies (a : b :: c : d), equivalence-equivalence requires matching same functional relations. Experiments 1 to 4 examined equivalence-equivalence in 5-year-old children. In each experiment, subjects were tested for equivalence-equivalence before equivalence and, if they did not show equivalence-equivalence, also after the equivalence test. The experiments included various procedural arrangements designed to facilitate equivalence-equivalence, all of which failed. Only 8/18 children showed equivalence-equivalence, 2 before (11%) and 6 after equivalence (33%), irrespective of the facilitative procedures that were used. Adults served in Experiment 5. This experiment was the same as Experiments 1 through 4 but without facilitative arrangements. All adults showed equivalence-equivalence, most of them before equivalence. These and previously collected findings (Carpentier, Smeets, & Barnes-Holmes, 2002) suggest that equivalence-equivalence is an age-related performance similar to that which has been reported in earlier developmental studies on classical analogies. Yet, one should be cautious using equivalence-equivalence as a model for analogical reasoning. The testing procedures in both types of tasks are sufficiently different to permit the performances to be based on different behavioral processes.

Recent studies have shown that after being trained on multiple arbitrary match-to-sample tasks, humans match not only functionally same stimuli but also functionally same stimulus relations (Barnes, Hegarty, & Smeets, 1997; Stewart, Barnes-Holmes, Roche, & Smeets, 2001). In Experiment 1 of the Barnes et al. (1997) study, for example, 5 adults and a 12-year-old boy were

Requests for reprints may be addressed to Franck Carpentier (Carpentier@fsw.leidenuniv.nl) or Paul M. Smeets (Smeets@fsw.leidenuniv.nl), Department of Psychology, Leiden University, P.O. Box 9555, Leiden, The Netherlands.

trained on four A-B and four A-C relations. After passing the equivalence test (B-C, C-B), the subjects received tests with BC compounds as samples and as comparisons (BC-BC). This test, hereafter referred to as the equivalence-equivalence test, measured whether the subjects matched compounds with same-class elements to other compounds with same-class elements (e.g., B1C1-B3C3: equivalence-equivalence) and compounds with different-class elements to other compounds with different-class elements (e.g., B1C2-B3C4: nonequivalence-nonequivalence). All subjects demonstrated equivalence-equivalence and nonequivalence-nonequivalence. Similar findings were obtained when, in Experiment 2 of that study (5 adults and a 9-year-old boy), equivalence-equivalence was tested before equivalence.

Equivalence-equivalence may be an important phenomenon, not only because it extends equivalence (Barnes, 1994; Saunders & Green, 1992; Sidman, 1994, 2000; Sidman & Tailby, 1982) but also because of its apparent correspondence with classical analogy tasks. Consider the following classical analogy task: "apple is to banana as sheep is to fish or book?" (a : b :: c : d). If apple and banana are equivalent in the context of fruit, and sheep and fish are equivalent in the context of animals, a person would be expected to select fish rather than book. If the b-term had been toy, the subject would probably select book. Thus, in spite of the different assessment procedures, equivalence-equivalence and classical analogies both require subjects to match functionally same stimulus relations. Perhaps, therefore, equivalence-equivalence could be used as a behavior analytic model for analyzing analogical reasoning.

Carpentier, Smeets, and Barnes-Holmes (2002) examined if the emergence of equivalence-equivalence follows the same developmental trend as analogical reasoning. Earlier developmental studies reported that analogical competence is rarely if ever found before the stage of formal operations (at about 12 years or later) and that younger children, especially before 9, have difficulty solving even the simplest analogy task (Levinson & Carpenter, 1974; Lunzer, 1965; Piaget, Montangero, & Billeter, 1977; Sternberg & Rifkin, 1979). More recent studies, however, reported that children as young as 3 years and even monkeys are capable of solving analogies (Alexander et al., 1989; Bovet & Vauclair, 2001; Goswami, 1991; Goswami & Brown, 1989, 1990; Thompson & Oden, 2000). The Carpentier et al. (2002) study consisted of four experiments. Experiment 1 was a straightforward replication of Experiment 1 of the Barnes et al. (1997) study with adults and with 9- and 5-year-old children. Experiments 2 and 3 were variations of Experiment 1 but with the same training and testing sequence: Training and testing baseline (A-B, A-C), testing equivalence (B-A, C-A, and B-C, C-B), and testing equivalence-equivalence (BC-BC). Almost all adults (88%) and 9-year olds (88%) but none of the 5-year-olds (0%) passed the equivalence-equivalence test. Thus, in as far as equivalence-equivalence and classical analogies are based on the same process (i.e., matching functionally same relations), these findings are consistent with the developmental studies which indicated that analogical competence is a late rather than early childhood phenomenon. However, all 5-year-olds did pass the equivalence-equivalence test when, in Experiment 4, they were first given the opportunity to relate baseline compounds with same-class elements to other compounds with same-class elements (e.g., A1B1-A3B3, A2C2-A1C1) and baseline compounds with different-class elements to other compounds with different-class elements (A2B3-A2B1; A3C1-A3C2) (baseline-baseline test). Thus, like with equivalence (Adams, Fields, & Verhave, 1993; Fields et al., 2000), equivalence-equivalence performances may be enhanced by the prior exposure to easier and/or prerequisite tasks.

The present study examined if 5-year-old children will, albeit with the help of extra training and testing arrangements, also show equivalence-equivalence *before* equivalence (Barnes et al., 1997, Experiment 2). The study consists of five experiments. Experiment 1 is a replication of Experiment 4 of the Carpentier et al. (2002) study except that now equivalence-equivalence was measured before, and if necessary, also after equivalence. Experiments 2 to 5 examined if the negative findings of Experiment 1 could be related to inappropriate testing arrangements or age.

Experiment 1

Method

Subjects

Eight 5-year-old children served as subjects. The age and sex of the individual subjects are presented, together with the results, in Table 1. The subjects were recruited through school contacts and served, with the parents' approval, on a voluntary basis. According to the teachers, none of the subjects had participated in experimental research before.

Setting, Observers, and Sessions

The experiment was conducted in a quiet room of the school building. All subjects were exposed to the procedures individually. The experimenter and subject were seated at the same table facing one another. The experimenter had received extensive training on the correct execution of the procedures with special emphasis on the prevention of any cues that could influence the subjects' responses (Saunders & Williams, 1998). Four other adults served as reliability observers, one at a time. The observer was present in the same room but situated such (i.e., behind and slightly to the left or right of the subject) that he/she could clearly observe the subject's responses, but not the experimenter's data sheet. The subjects required 16 to 31 daily sessions of 16 to 22 min each over a period of 18 to 48 days.

Stimuli

The stimuli were nine black abstract forms (3 x 3 cm), (see Figure 1). For convenient reference, the stimuli are identified here by alphanumeric codes (e.g., A1, B2, B3). The forms were presented on laminated white cards (15 x 21 cm) and were presented as unitary stimuli (e.g., A1 or B1) or as compounds (e.g., B1C1). The compound elements were presented side by side. Additional materials included a subject registration form. This form consisted of a string of 100 cells numbered 1 to 100.

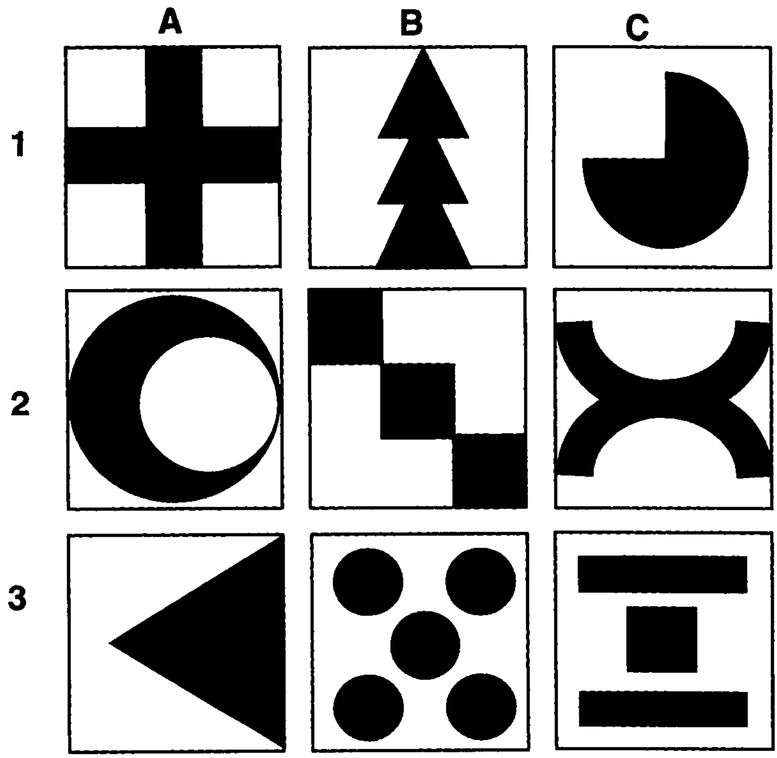

Figure 1. Stimuli used in Experiments 1, 2, and 5.

Tasks and Programmed Consequences

Simultaneous match-to-sample tasks were used for training and testing. Each stimulus card showed multiple horizontally aligned comparisons in the upper part, and a sample below.

During training, responses defined as correct were followed by verbal approval ("Right," "Correct," "Good boy/girl") and the experimenter marking the next cell on the subject's registration form. Immediately after receiving the 100th mark, the experimenter interrupted the session, permitted the subject to select a picture card of his/her choice (e.g., animal, flower, cartoon character), presented a new form, and continued with the following trial. Responses defined as incorrect were followed by negative feedback ("Wrong"). During testing, responses were followed by no programmed consequences other than the presentation of the next trial.

Test and Training Sequence

The program consisted of 12 phases. At first, the subjects were trained and tested on the A-B and A-C baseline tasks (Phases 1 through 4). Then they were trained to match the A-B and A-C sample-comparison relations to corresponding AB and AC compounds in Phase 5. This training was to ensure that (a) these young subjects attended to both elements of the AB and AC compounds, and (b) the sample-comparison relations were functionally equivalent to these compounds (e.g., A1-B1 = A1B1, A3-C2 = A3C2). In Phase 6, the subjects were given the opportunity to match baseline compounds with same-class elements and baseline compounds with different-class elements (e.g., A1B1-A3B3, A3C1-A3C2) (baseline-baseline test). If necessary, these relations were trained in Phase 7. After receiving further (sample-comparison)-compound training in Phase 8 (B-C = BC), the subjects were tested for equivalence-equivalence (BC-BC), symmetry (B-A, C-A), and equivalence (B-C, C-B) in Phases 9, 10, and 11, respectively. If necessary, responding according to equivalence was trained in Phase 12. Subjects who had failed the equivalence-equivalence test before having shown equivalence, now received the equivalence-equivalence test again.

Phase 1: Training A-B. Three relations were trained, A1-B1, A2-B2, and A3-B3. These tasks were trained in a multiple-step fashion. A1-B1 and A2-B2 were trained in Step 1. Blocks of 18 trials were used, 9 A1-B1 trials quasi-randomly mixed with 9 A2-B2 trials. On each trial, the experimenter presented a card while saying, "Point." Each block was preceded by two demonstration trials, one on each task. On each demonstration trial, the experimenter (a) pointed to the sample while saying, "Look here," (b) pointed to the correct comparison while saying, "I point to this," and (c) invited the subject to point to the same comparison, "Now you point." Training continued until a subject responded correctly on all 18 trials of one block or on 33/36 trials of two consecutive blocks with no more than two errors on a same task (e.g., A1-B1).

In Step 2, A3-B3 was added. Each block started with an A3-B3 demonstration trial, followed by 18 no-help trials, 9 A3-B3 trials (50%) mixed with 4 A1-B1 and 5 A2-B2 trials (50%) or 5 A1-B1 and 4 A2-B2 trials (50%). Subjects who demonstrated criterion performance (same as in Step 1) proceeded to the next step. Step 3 was the same as Step 2 except that no demonstration trials were used and each block involved six trials on each A-B task.

Phases 2 and 3: Training A-C and A-B. These phases were directed at establishing the A-C relations (A1-C1, A2-C2, A3-C3) while maintaining the A-B performances. The A-C relations were trained in Phase 2. The procedures were the same as for the A-B training (three steps) except that each block of 18 A-C trials was followed by a block of 6 A-B trials. In each step, training continued until a subject responded correctly on all 18 A-C trials and on 5/6 A-B trials, or on 33/36 A-C trials with no more than two errors on the same task and on 11/12 A-B trials. In Phase 3, A-B and A-C were trained in a mixed fashion. Blocks of 18 trials were used, 9 A-B trials quasi-randomly interspersed among 9 A-C trials.

Phase 4: Testing A-B and A-C. This phase assessed if the trained performances remained intact under testing conditions. Two test blocks were used, one of 18 A-B trials and one of 18 A-C trials. Each test block was followed by a block of 6 training trials, 3 A-B trials mixed with 3 A-C trials (Block 1: Test A-B; Block 2: Train A-B and A-C; Block 3: Test A-C; Block 4: Train A-B and A-C). Immediately before each *test* block, the experimenter eliminated the subject's registration form and said, "Now we are going to play the game without me telling you whether you are right or wrong. I will also not give you any marks. Later you can earn marks again." From that moment on, the experimenter refrained from any communication and silently presented one trial after another. Immediately before each *training* block, the experimenter placed the subject's registration form on the table while saying, "Now you can earn marks again." Subjects who demonstrated criterion performance (same as in Phases 2 and 3) advanced to Phase 5.

Phase 5: (Sample-comparison)-compound training ([A-B]-AB and [A-C]-AC). The subjects were trained to relate A-B and A-C sample-comparison relations to corresponding AB and AC compounds which, in Phase 6, were used for measuring baseline-baseline (AB-AB, AC-AC) in Phases 6 and 7. Blocks of 18 trials were used. Nine of these trials involved same-class sample-comparison relations (e.g., [A1-B1]-A1B1, [A3-C3]-A3C3). These trials were quasi-randomly mixed with nine trials involving different-class sample-comparison relations (e.g., [A1-B2]-A1B2, [A2-C3]-A2C3). Some of these trials are illustrated in Figure 2.

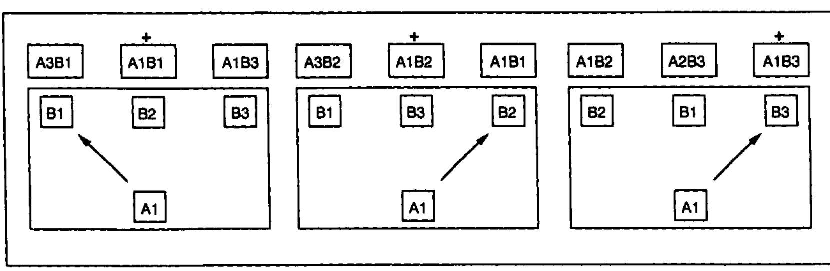

Figure 2. Examples of (sample-comparison)-compound tasks.

On each (A-B)-AB trial, the experimenter presented a stimulus card that was also used for baseline training and testing (e.g., A1 as sample and B1, B2, and B3 as comparisons). Immediately above that card, three AB compounds were shown (e.g., A3B1, A1B1, A1B3), one above each comparison. These compounds varied from trial to trial. Each trial required the subjects to attend to the demonstrated sample and comparison, and to both elements of each compound. The locations of the correct compounds (left, center, right) varied unsystematically over trials. The same procedures were used for the (A-C)-AC trials. Each block was preceded by two demonstration trials. On each demonstration trial, the experimenter gave the following instruction, "This (while pointing to the sample and designated comparison) goes with this" (while pointing to the designated compound), and invited the subject to point to the same compound ("Point here"). Subjects who responded correctly on at least 17/18 trials proceeded to Phase 6.

Phases 6 and 7: Testing and training baseline-baseline (AB-AB, AC-AC). Phase 6 (baseline-baseline test) measured whether the subjects matched baseline compounds with same-class elements to other compounds with same-class elements and baseline compounds with different-class elements to other compounds with different-class elements (e.g., A1B1-A3B3, A2C3-A2C1). Figure 3 shows schematic illustrations of some of these trials. The procedures were the same as during the baseline test (i.e., two blocks of test trials and two blocks of training trials). Each test block consisted of six AB-AB trials mixed with six AC-AC trials. If necessary, this test was presented two more times. Subjects who responded correctly on at least 22/24 test trials and on 11/12 training trials proceeded directly to Phase 8. Those who consistently failed this test were trained on these tasks in Phase 7. The training procedures were the same as for testing (Phase 6) except that feedback was provided on all trials. Training continued until criterion was reached (same as in Phase 6) or 72 trials had been completed. At that point, the subjects advanced to the next phase, irrespective of their performance.

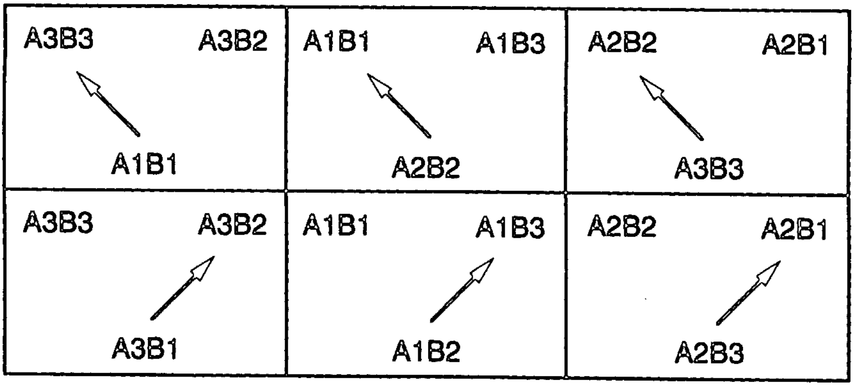

Figure 3. Examples of baseline-baseline tasks.

Phase 8: (Sample-comparison) compound training ([B-C]-BC). The subjects were trained to relate the B-C sample-comparison relations to corresponding BC compounds that were later used for measuring equivalence-equivalence in Phase 9. Blocks of 18 trials were used, 9 trials with equivalent sample-comparison relations (e.g., [B1-C1]-B1C1) mixed with 9 trials with nonequivalent sample-comparison relations (e.g., [B1-C2]-B1C2). The procedures were the same as in Phase 5 (training [A-B]-AB, [A-C]-AC).

Phase 9: Testing equivalence-equivalence (BC-BC). This test measured whether the subjects matched BC compounds with same-class elements to other BC compounds with same-class elements (equivalence-equivalence), and BC compounds with different-class elements to other BC compounds with different-class elements (see Figure 4). In the interest of brevity, relating equivalence to equivalence and nonequivalence to nonequivalence will be described here as equivalence-equivalence responding. The procedures were the same as for the baseline-baseline test (Phase 6). Each test block consisted of 16 BC-BC trials: 4 B1C1-B3C3, 4 B1C2-B2C3, 4 B2C2-B3C3, and 4 B2C1-B1C3 trials. Each test block was followed by a block of six baseline training trials (A-B and A-C). Subjects who responded correctly on at least 28/32 test trials (no more than two errors on the same task) and on 11/12 training trials proceeded to Step 10 (symmetry testing). Subjects who did not show equivalence-equivalence received a maximum of two more presentations of this test. After the third presentation, the subjects advanced to the next phase irrespective of their equivalence-equivalence performance.

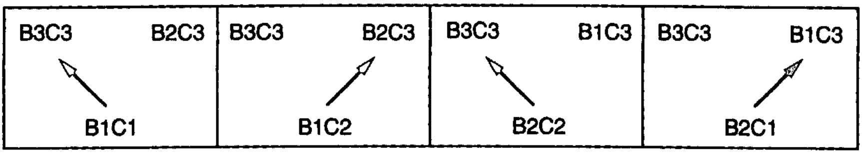

Figure 4. Equivalence-equivalence tasks.

Phases 10, 11, and 12: Testing symmetry (B-A, C-A), testing and training equivalence (B-C, C-B). The procedures were the same as in Phase 4 (i.e., two blocks of 18 test trials, each followed by a block of six baseline training trials). Symmetry (B-A, C-A) was tested in Phase 10, and equivalence (B-C, C-B) in Phase 11. The subjects received a maximum of three successive presentations of each test. Subjects who consistently failed the equivalence test received training on these tasks in Phase 12 (same procedures as for the baseline-baseline training, see Phase 7). Subjects who had failed the equivalence-equivalence test before equivalence now again received the equivalence-equivalence test (maximum of two more presentations). Each test failure was followed by a presentation of the equivalence test to assess if the B-C and C-B relations were still intact.

Interobserver Reliability

1717 trials (25%) were monitored. A trial was considered an agreement if the experimenter and observer recorded the same response. The experimenter and observers agreed on all training trials (100%) and on all but four test trials (99%).

Results and Discussion

The results are shown in Table 1. The training results are expressed in numbers of trials for reaching criterion. Test results are expressed in terms of pass (+) or fail (-). All subjects learned the baseline tasks (188-240 trials), continued to respond accurately on these tasks whenever presented during subsequent testing and training conditions, and related demonstrated sample-compound relations to corresponding compounds ([A-B]-AB, [A-C]-AC, [B-C]-BC).

Table 1

Training and Test Results in Experiment 1 (Children)

		Subjects								
Tasks	Train Test	No Age Sex	1 5;3 M	2 5;4 F	3 5;4 M	4 5;5 M	5 5;5 M	6 5;5 F	7 5;7 F	8 5;9 F
A-B, A-C	Train		214	188	208	212	188	232	240	230
A-B, A-C	Test		+	+	+	+	+	+	+	+
(A-B)-AB, (A-C)-AC	Train		18	18	18	18	18	18	18	18
AB-AB, AC-AC	Test		-	-	-	-	-	-	-	-
			-	-	-	-	-	-	+	-
			-	+	-	+	+	-		-
AB-AB, AC-AC	Train		72*		72*			72*		72*
(B-C)-BC	Train		18	18	18	18	18	18	18	18
BC-BC	Test		-	-	-	-	-	-	-	-
			-	-	-	+	-	-	-	-
			-	-	-		-	-	-	-
B-A, C-A	Test		-	+	+	+	+	+	+	+
			-							
			+							
B-C, C-B	Test		-	-	-	+	+	+	+	+
			+	-	-					
				-	-					
B-C, C-B	Train			36	54					
BC-BC	Test		-	-	-		-	-	+	-
B-C, C-B	Test		+	+	+		+	+		+
BC-BC	Test		-	-	-		+	-		-

Note. M = male, F = female. Training results are expressed in terms of required numbers of trials. Tests results are expressed in terms of pass (+) or fail (-). (*) = training failed.

Only 4 subjects (2, 4, 5, and 7) passed the baseline-baseline test. The other 4 subjects failed to demonstrate these relations during testing *and* training. Six subjects (1, 4, 5, 6, 7, and 8) demonstrated equivalence. Of these subjects, 3 demonstrated equivalence-equivalence, 1 before (4), and 2 after equivalence (5 and 7). These 3 subjects had passed the baseline-baseline test before. All 3 subjects who failed the equivalence-equivalence test (1, 6, and 8) had also failed the baseline-baseline test.

Thus, consistent with our previous findings (Carpentier et al., 2002, Experiment 4), subjects who passed the baseline-baseline also passed the equivalence-equivalence test. Contrary to the previous study, however, was the high proportion of subjects (50%) failing the baseline-baseline test. Perhaps, for these 4 children, the transition from baseline (three unitary stimuli) to baseline-baseline (three 2-element compounds) had been too abrupt.

Experiment 2

This experiment evaluated a program designed to facilitate the transition from the baseline to baseline-baseline. The program, hereafter referred to as the FACES-program, consisted of multiple steps. This program started with a (pre)test, in which the AB and AC compounds served as samples and two schematic face drawings as comparisons, a HAPPY FACE and a SAD FACE (see Figure 5). These stimuli were selected on the assumption that for most 5-year-olds HAPPY FACE and SAD FACE are functionally equivalent with correct and incorrect, respectively. Would subjects relate compounds with same-class elements to HAPPY FACE and compounds with different-class stimuli to SAD FACE (e.g., A1B1&A3B3-HAPPY FACE and A2B3&A2C1-SAD FACE)? If not, the subjects were trained on a subset of these tasks and retested on all (trained and untrained) tasks. If necessary, this test and training cycle continued until all tasks were trained. Note that this program (a) uses stimulus configurations (compound samples and unitary comparisons) that are more complex than those in the baseline tasks (all unitary stimuli) and are less complex than those in the baseline-baseline tasks (three 2-element compounds), (b) involves multiple-exemplar training that has been effectively used in previous research on emergent performances (Baer, Peterson, & Sherman, 1967; Barnes-Holmes, Barnes-Holmes, Roche, & Smeets, 2001; Cullinan, Barnes-Holmes, & Smeets, 2001; Schusterman & Kastak, 1993; Stokes & Baer, 1977), and (c) permits assessment of class formation (e.g., A1B1&A3B3-HAPPY FACE, A1B2&A2B3-SAD FACE). Would most subjects pass this test? If so, would these subjects also pass the more complex baseline-baseline test (e.g., A1B1&A3B3-HAPPY FACE, hence A1B1-A3B3; A1B2&A2B3-SAD FACE, hence A1B2-A2B3) and, perhaps as a result thereof, also show equivalence-equivalence?

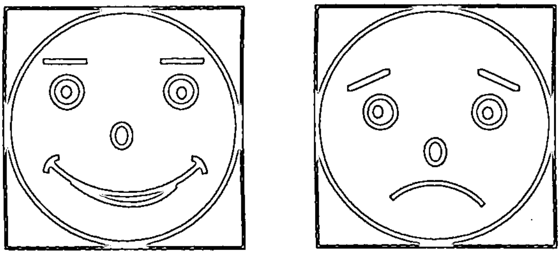

Figure 5. Happy Face and Sad Face.

Method

Four new 5-year-old children served as subjects (see Table 2). The recruitment procedures, setting, stimulus materials, and programmed consequences were the same as in Experiment 1. The same applied to the training and testing sequence except that the FACES-program was added (see below). This program was introduced immediately before the baseline-baseline test. The experiment required 24 to 36 sessions (per subject) of about 15 min each during a period of 51 to 68 days.

FACES-Program: AB&AC-FACE

Test. Two test blocks of 12 trials each were used. Each test block was followed by a block of six baseline training trials. The first test block measured AB-FACE. Six of these trials involved the three compounds with same-class elements (A1B1, A2B2, or A3B3), two trials each. These trials were quasi-randomly mixed with six trials involving compounds with different-class elements (A1B2, A1B3, A2B1, A2B3, A3B1, or A3B2), one trial each. The second test block measured AC-FACE (same procedure). A response was scored correct when a subject pointed to HAPPY FACE when given a compound with same-class elements, and to SAD FACE when given a compound with different-class elements. The test was presented without any introduction. Subjects who responded correctly on at least 22/24 test trials and on 11/12 training trials, proceeded to the baseline-baseline test (AB-AB, AC-AC). Those who failed the FACE test were trained on a subset or all of these trials (see below). All subjects received the test at least once and maximally four times (i.e., [pre]Test → Training Subset 1 → Test → Training Subset 2 → Test → Training Subset 3 → Test).

Training Subset 1. The subjects were trained to relate two compounds with same-class elements to HAPPY FACE, and four compounds with different-class elements to SAD FACE. Blocks of 16 training trials were used. Each block included 8 trials with A1B1 and A2C2 as samples (4 trials each). These trials were quasi-randomly mixed with 8 trials with A1B3, A3B1, A1C2, and A2C1 as samples (2 trials each). Each block was preceded by 4 demonstration trials with samples A1B1, A2C2, A1B3, or A2C1. During each demonstration trial, the experimenter first pointed to the sample, then to the designated face, and finally gave the subject the opportunity to do the same. Subjects who responded correctly on at least 30/32 trials, received the test again.

Training Subsets 2 and 3 was the same except that different samples were used: A3B3, A1C1, A2B3, A3B2, A1C3, A3C1 (Subset 2), and A2B2, A3C3, A1B2, A2B1, A2C3, A3C2 (Subset 3).

Interobserver Reliability

711 trials were monitored (17%). The experimenter and observers agreed on all but two trials, one training trial and a test trial.

Results and Discussion

The results are shown in Table 2. All subjects completed the baseline training successfully (188-257 trials), continued to respond accurately on these tasks during subsequent testing and training conditions, and successfully completed all (sample-comparison)-compound tests.

Table 2

Training and Test Results in Experiment 2 (Children)

			Subjects			
		No	9	10	11	12
	Train	Age	5;0	5;0	5;1	5;1
Tasks	Test	Sex	M	M	F	F
A-B, A-C	Train		225	188	225	257
A-B, A-C	Test		+	+	+	+
(A-B)-AB, (A-C)-AC	Train		18	18	18	18
AB&AC-Face	Test		-	+	-	-
AB&AC-Face Subset 1	Train		128		128	64
AB&AC-Face	Test		-		-	+
AB&AC-Face Subset 2	Train		40		40	
AB&AC-Face	Test		+		-	
AB&AC-Face Subset 3	Train				96	
AB&AC-Face	Test				+	
AB-AB, AC-AC	Test		+	-	-	-
				-	-	+
				+	-	
AB-AB, AC-AC	Train				48	
(B-C)-BC	Train		18	18	18	18
BC-BC	Test		-	-	-	-
			+	-	-	-
				-	-	-
B-A, C-A	Test		+	+	+	+
B-C, C-B	Test		+	+	+	-
						-
						+
BC-BC	Test			-	-	-
B-C, C-B	Test			+	+	+
BC-BC	Test			-	-	-
B-C, C-B	Test			+	+	+
BC-BC	Test					-
B-C, C-B	Test					+

Note. Notations are as in Table 1.

Three subjects related the AB and AC compounds with same-class elements to HAPPY FACE and those with different-class elements to SAD FACE, Subject 10 already during the pretest, and Subjects 12 and 9 after being trained on one and two subsets. These 3 subjects also passed the baseline-baseline test. Subject 11 required training on all AB&AC-FACE tasks (i.e., no evidence of class formation). She also failed the baseline-baseline test but quickly learned these tasks during training. Only Subject 9 demonstrated equivalence-equivalence before equivalence. The other 3 subjects consistently failed the equivalence-equivalence tests before and after demonstrating equivalence.

In conclusion, most subjects responded class-consistently during the FACES-program (AB&AC-FACE) and during the baseline-baseline test (A1B1&A3B3-HAPPY FACE, hence A1B1-A3B3). The proportion of subjects showing equivalence-equivalence (25%), however, was even lower than in Experiment 1 (38%).

Experiment 3

This experiment examined if equivalence-equivalence could be facilitated by using familiar stimuli. Holth and Arntzen (1998) reported that equivalence emerges more readily when using familiar and easy-to-name stimuli (e.g., chair, cowboy) rather than with abstract stimuli (Greek letters). Perhaps, this also applies to equivalence-equivalence.

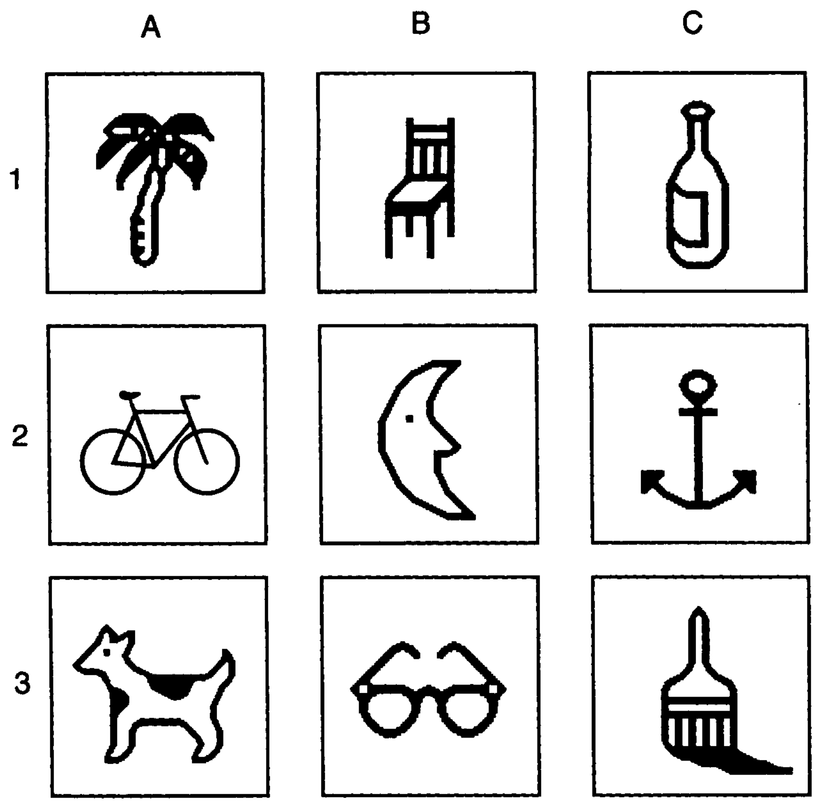

Figure 6. Stimuli used in Experiments 3 and 4.

Method

Four new children participated (see Table 3). The procedures were the same as in Experiment 2 except that now familiar stimuli were used (see Figure 6). The children were screened on their ability to name these stimuli. The program required 19 to 26 sessions (per subject) of about 17 min each during a period of 29 to 43 days. 1041 trials (27%) were monitored. The experimenter and observers agreed on all trials.

Results and Discussion

All subjects learned the baseline tasks (168-232 trials), continued to respond accurately on these tasks during subsequent training and testing conditions, and made no or very few errors on the (sample-comparison)-compound tasks.

Table 3

Training and Test Results in Experiment 3 (Children)

		Subjects				
Tasks	Train Test	No Age Sex	13 5;7 M	14 5;7 M	15 5;9 F	16 5;9 M
A-B, A-C	Train		194	232	168	188
A-B, A-C	Test		+	+	+	+
(A-B)-AB, (A-C)-AC	Train		18	18	18	18
AB&AC-Face	Test		+	-	-	-
AB&AC-Face Subset 1	Train			64	64	128
AB&AC-Face	Test			+	+	-
AB&AC-Face Subset 2	Train					64
AB&AC-Face	Test					+
AB-AB, AC-AC	Test		-	-	-	-
			+	-	-	+
				-	-	
AB-AB, AC-AC	Train			56	28	
(B-C)-BC	Train		18	18	18	18
BC-BC	Test		-	-	-	-
			-	-	-	-
			-	-	-	-
B-A, C-A	Test		+	+	+	+
B-C, C-B	Test		+	+	+	+
BC-BC	Test		-	+	-	-
B-C, C-B	Test		+		+	+
BC-BC	Test		+		-	-
B-C, C-B	Test				+	+

Note. Notations are as in Table 1.

All 4 subjects passed the AB&AC-FACE test, Subject 13 immediately (pretest), and Subjects 14, 15, and 16 after being trained on one or two subsets. Two subjects (13, 16) also passed the baseline-baseline test. Both subjects who failed the baseline-baseline test (14, 15) learned these tasks through training. All 4 subjects demonstrated equivalence, 2 of whom also showed equivalence-equivalence, both after equivalence: Subject 13 who had passed the baseline-baseline test before, and Subject 14 who had failed that test. These findings were very similar to those obtained in Experiments 1 and 2. Stimulus familiarity, therefore, did not facilitate equivalence-equivalence.

Experiment 4

In Experiments 1 to 3, only 9/16 children (56%) passed the baseline-baseline test. Of the 14 children who demonstrated equivalence, only 6 (43%) showed equivalence-equivalence, 2 before and 4 after equivalence. These findings suggested that, for many of these children, the compound-compound matching tasks were too difficult, probably because of the complexity of the stimulus configurations. Experiment 4, therefore examined how young children would respond on a less complex BC-FACE test (e.g., B1C1&B3C3-HAPPY FACE, B2C1&B1C3-SAD FACE). Although this test does not measure equivalence-equivalence (because it does not require subjects to match compounds on the basis of the equivalent or nonequivalent relations between elements), it could be seen as the best possible approximation thereof. Specifically, the following questions were addressed: Will most subjects showing AB&AC-FACE also show BC-FACE? If so, will they pass that test before or after demonstrating equivalence? Finally, will most subjects showing BC-FACE also show equivalence-equivalence (e.g., B1C1&B3C3-HAPPY FACE, hence B1C1-B3C3; B2C1&B1C3-SAD FACE, hence B2C1-B1C3)?

Method

Four new children participated (see Table 4). The procedures and stimuli were the same as in Experiment 3 except that (a) the baseline-baseline test (AB-AB, AC-AC) was no longer used, and (b) the equivalence-equivalence test (BC-BC) was preceded by a BC-FACE test. Briefly, the test and training sequence was as follows. After having learned the baseline tasks (A-B, A-C) and to match these sample-comparison relations with corresponding compounds ([A-B]-AB, [A-C]-AC), the subjects received the FACE-program (AB&AC-FACE). After also learning to relate B-C sample-comparison relations to corresponding BC compounds (e.g., [B1-C1]-B1C1, [B3-C2]-B3C2), the BC-FACE test was introduced.

BC-FACE Test. Two test blocks of 18 trials were used. Each block consisted of 9 BC-FACE trials, 4 BC-HAPPY FACE trials quasi-randomly mixed with 5 BC-SAD FACE trials (Block 1) or 5 BC-HAPPY FACE trials mixed with 4 BC-SAD FACE trials (Block 2). These trials were quasi-randomly mixed with 9 AB&AC-FACE trials. Each test block was followed by a block of 6 baseline-training trials (A-B, AC). The procedures were the same as for the AB&AC-FACE test (Experiments 1 and 3), except that no demonstration trials were used. Criterion was set at 32 or more trials correct (i.e., at least 16/18 BC-FACE and 16/18 AB&AC-FACE trials correct). If necessary, the test was presented two more times in succession. Subjects who passed the test proceeded to the equivalence-equivalence test. Those who failed the BC-FACE test proceeded to the symmetry and equivalence tests before receiving the BC-FACE test again. If they passed the BC-FACE test now, they also received the equivalence-equivalence test.

The program required 23 to 28 daily sessions (per subject) of about 17 min each, during a period of 37 to 44 days. 988/3852 trials were monitored (26%). The experimenter and observers agreed on all but one (test) trial.

Table 4

Training and Test Results in Experiment 4 (Children)

			Subjects			
		No	17	18	19	20
	Train	Age	5;1	5;2	5;2	5;4
Tasks	Test	Sex	F	F	M	M
A-B, A-C	Train		188	266	212	188
A-B, A-C	Test		+	+	+	+
(A-B)-AB, (A-C)-AC	Train		18	18	18	18
AB&AC-Face	Test		-	-	-	-
AB&AC-Face Subset 1	Train		112	64	32	32
AB&AC-Face	Test		+	+	+	+
(B-C)-BC	Train		18	18	18	18
BC-Face*	Test		+	-	-	-
				-	-	-
				-	-	-
BC-BC	Test		-			
BC-Face*	Test		+			
BC-BC	Test		-			
BC-Face*	Test		+			
BC-BC	Test		-			
B-A, C-A	Test		+	+	+	+
B-C, C-B	Test		+	+	+	+
BC-Face*	Test			+	-	+
					+	
BC-BC	Test		+	-	-	-
BC-Face*	Test			+	+	+
BC-BC	Test			-	-	+
BC-Face*	Test			+	+	
BC-BC	Test			-	-	

Note. Notations are as in Table 1.

Results and Discussion

All subjects completed the baseline training successfully (188-266 trials), continued to respond accurately on these trials during subsequent test and training phases, related demonstrated A-B, A-C, and B-C sample-comparison relations to corresponding compounds, and evidenced class formation of trained compounds (AB&AC-FACE) after being trained on one subset (See Table 4). All 4 subjects demonstrated the designated BC-FACE relations, 1 before (17) and 3 after equivalence (18, 19, 20). Two of these subjects (17, 20) also showed equivalence-equivalence, both after equivalence.

In conclusion, all 4 subjects evidenced BC-FACE performances indicative of class-extension (i.e., from AB&AC-FACE to AB&AC&BC-FACE) though most of them after equivalence. The proportion of equivalence-equivalence occurrences (50%) was about the same as in Experiments 1 to 3 (43%). BC-FACE, therefore, did not facilitate equivalence-equivalence.

Experiment 5

This experiment examined if the many equivalence-equivalence failures in Experiments 1 to 4, notably those before equivalence, were related to the subjects' age or to some difficult-to-identify procedural inadequacies. Would our current procedures also prevent adults from showing equivalence-equivalence? If not, would most of these subjects, like in the study by Barnes et al. (1997), show equivalence-equivalence before equivalence?

Method

Four adults served as subjects (see Table 5), a primary-school teacher, a first-year psychology student, and two law students. The subjects were recruited through a bulletin board announcement and were paid for their participation. The experiment was conducted in a laboratory room of the Psychology Department at Leiden University. All subjects received an abbreviated version of the testing and training program that was used in Experiment 1. Following the training and testing of the baseline tasks (A-B, A-C), the subjects proceeded directly to the equivalence-equivalence test (BC-BC). After also demonstrating equivalence, subjects who had failed the equivalence-equivalence test before, now received that test again (same sequence as in Experiment 2 of the Barnes et al., 1997, study). The stimuli and procedures for testing and training were the same as in Experiment 1 except that no subject registration forms and no back-up reinforcers were used (only verbal feedback). The subjects completed the experiment in one or two sessions, each lasting 75 to 95 min. 480 trials (27%) were monitored. The observers and experimenter agreed on all trials.

Results and Discussion

All 4 adults learned the baseline tasks in 168 trials (no range), continued to respond accurately on these tasks during subsequent testing and training conditions, and demonstrated equivalence-equivalence, 3 before (21, 23, 24), and 1 (22) after equivalence (see Table 5). These findings are consistent with those reported by Barnes et al. (1997, Experiment 2). Thus, the equivalence-equivalence failures in Experiments 1 to 4 of the current study probably related to the subjects' age and/or school history.

Table 5

Training and Test Results in Experiment 5 (Adults)

			Subjects			
		No	21	22	23	24
	Train	Age	21	23	26	45
Tasks	Test	Sex	F	F	F	F
A-B, A-C	Train		168	168	168	168
A-B, A-C	Test		+	+	+	+
BC-BC	Test		-	-	-	-
			+	-	+	-
				-		+
B-A, C-A	Test		+	+	+	+
B-C, C-B	Test		+	+	+	+
BC-BC	Test			+		

Note. Notations as in Table 1.

General Discussion

Present findings are consistent with those reported before (Carpentier et al., 2002) in that equivalence-equivalence appears to be an age-related phenomenon. In the previous study (Carpentier et al., 2002), when equivalence-equivalence was measured exclusively after equivalence, 7/8 adults (88%), 7/8 9-year-old children (88%), and 4/16 5-year-old children (25%) showed equivalence-equivalence. These proportions are very similar to those obtained in the present study: equivalence-equivalence in all 4 adults (100%) and only in 8/18 children (44%) who demonstrated equivalence (Fisher test, p = .067). This discrepancy becomes even more pronounced when considering that (a) the children's programs included many additional training and testing arrangements that were designed to facilitate equivalence-equivalence responding, and (b) only 2/18 children (11%) as opposed to 3/4 adults (75%) showed equivalence-equivalence before equivalence (Fisher test, p = .023). Thus, young children show equivalence-equivalence not only less often than older subjects, but whenever they do, also less readily.

In so far as equivalence-equivalence and classical analogies are based on the same process (i.e., matching functionally same stimulus relations), these findings are consistent with the earlier developmental studies which indicate that analogical competence is not an early childhood phenomenon (Gallagher & Wright, 1979; Levinson & Carpenter, 1974; Sternberg & Rifkin, 1979). On balance, the proportions of 5-year-olds showing equivalence-equivalence are sufficiently high for lending support to later developmental studies reporting that analogical competence occurs in children much younger than 12 years (Alexander et al.,1989; Goswami, 1991; Goswami & Brown, 1989, 1990).

This age effect could be related to the complexity of the stimulus configurations. The failures to obtain class-consistent compound-compound performances from compound-FACE relations (e.g., A1B1&A3B3-HAPPY FACE, yet no A1B1-A3B3, and B1C1&B2C2-SAD FACE, yet no B1C1-B2C2) in Experiments 2 to 4, and the failures to establish the baseline-baseline tasks (AB-AB, AC-AC) through direct training in Experiment 1 pointed to this direction. This "complexity" account could be questioned given the (near) errorless performances during the (sample-comparison)-compound tasks (e.g., [A-B]-AB). These tasks required the subjects to attend to the demonstrated sample-comparison relation (e.g., A1-B1) and to all elements of three simultaneously presented compounds (A1B1, A2B1, A1B2). Although these stimulus configurations were not less complex than those used in the compound-compound tasks, they may have been easier because they required identity rather than arbitrary matching.

The equivalence-equivalence failures could also be related to the fact that these children had not yet been exposed to a formal academic (elementary school) program. These programs typically provide a multitude of activities fostering the matching of functionally same relations. For example, when introducing novel vocabulary such as "shelter," "communication," or "predator," teachers typically first provide a definition of the term (e.g., shelter is a place to hide or to relax), then give an example (bird-nest) and finally ask the students to mention the shelters of dogs, horses, babies, mice, and cars. Although these activities are not formatted as equivalence-equivalence (BC-BC) or as classical analogy tasks (a : b :: c : d), they may encourage students to relate functionally same stimulus relations with one another (bird-nest = horse-stable = car-garage). Tasks more directly related to our format equivalence-equivalence test can be found in basic math problems such as "1 pencil costs 5 cents. 3 pencils cost ....?" and "1/2 = ?/6. Perhaps, therefore, 5-year-olds could also acquire equivalence-equivalence through multiple-exemplar training (Barnes-Holmes et al., 2001; Cullinan et al., 2001; Schusterman & Kastak, 1993).

Given the different testing procedures, however, one should be cautious in using equivalence-equivalence as a model for classical analogies. First, not all classical analogies require equivalence-equivalence. Consider classical analogy tasks used by Goswami and Brown (1990) with young children such as: bird is to nest, as dog is to kennel, bone, cat, or other dog. Note that the relations between the a- and b-term and between the c-term and each d-term option have been directly trained or observed. Children have been told about or seen birds in nests, dogs in kennels, dogs chewing bones, and dogs interacting with cats and other dogs. In fact, had they not, they should be unable to solve these tasks (Goswami, 1991). The above type of analogy task, therefore, requires children to match stimulus pairs with same *trained* relations; hence it is similar to our baseline-baseline test (e.g., AB-AB). This measurement, however, would not change the outcome of the current study since, in Experiments 1 to 3, only 9/16 children passed the baseline-baseline test. This proportion is not much higher than the proportion of children (6/14) who, in the same experiments, passed the equivalence-equivalence test before or after equivalence.

Second, equivalence-equivalence and classical analogies differ with regard to the d-term options that are used. In the analogy tasks used by Goswami and Brown (1990), each d-term option fits or "goes with" the c-term, irrespective of the a- and b-terms. Each d-term option (i.e., kennel, bone, cat, and other dog) would be correct when given only the c-term (dog). Unlike equivalence-equivalence, therefore, classical analogy tasks do not require subjects to identify which stimulus-stimulus relations are correct or incorrect. Instead, they require subjects to choose out of several "correct" c-d relations (dog-kennel, dog-bone, etc.) the one that is functionally the same as the given a-b relation (bird-nest).

In fact, the emergent compound-compound matching performances could be seen as demonstrations of generalized matching stimuli with same discriminative functions of which the boundaries were set by the complexity of the stimulus configurations. Consider the performances of the 5-year-olds first. The baseline training (A-B, A-C) produced separable discriminative AB and AC compounds (Carpentier, Smeets, & Barnes-Holmes, 2000; Darcheville, Legrand, & Smeets, 1998; Dougher & Markham, 1994; Stromer, McIlvane, & Serna, 1993). The baseline-baseline tests allowed the subjects to match correct compounds with other correct compounds (e.g., A1B1-A3B3) and incorrect compounds with other incorrect compounds (e.g., A3B1-A3B2) which some children did and others did not, presumably because of the complexity of the stimulus configurations. From the equivalence test on, the situation repeated itself. B-C and C-B matching led to the formation of discriminative BC compounds (e.g., B1C1+ and B1C2-) so that, when given the equivalence-equivalence test, the subjects again matched compounds with same discriminative functions (e.g., B1C1-B3C3, and B1C2-B2C3) provided they attended to all compound elements.

This account also fits the emergent compound-FACE performances in Experiments 2 through 4. After completing the baseline training (A-B, A-C), the subjects were given the opportunity to relate the discriminative AB and AC compounds to unitary stimuli presumably with same discriminative properties (HAPPY FACE and SAD FACE). For some children, the discriminative properties of these stimuli existed from the onset, and they immediately related correct with correct (e.g., A1B1-HAPPY FACE) and incorrect with incorrect (e.g., A1B2-SAD FACE). For other children, the S+ and S- properties of HAPPY FACE and SAD FACE came about through transfer training (e.g., A1B1[+]&A3B3[+]-HAPPY FACE, hence HAPPY FACE [+]). Thus, from that point on, they did as before and matched compounds with faces of same discriminative properties, first the AB and AC compounds that had not been used for transfer training, and later (i.e., after demonstrating equivalence) the BC compounds.

The same may have occurred with the adults but without the exposure to the equivalence test. Training of A1-B1 and A1-C1 produced compounds A1B1+ and A1C1+, hence B1C1+; training of A3-B3 and A3-C3 produced compounds A3B3 and A3C3, hence B3C3+. During the equivalence-equivalence test, the subjects simply matched BC compounds with same discriminative functions, B1C1+ with B3C3+ and B1C2- with B2C3-. Although this account leaves the emergent discriminative BC compounds to be explained, previous research has shown that, after being trained on two AB and two AC simple discrimination tasks, (A1B1+/A1B2-, A2B2+/A2B1-, A1C1+/A1C2-, A2C2+/A2C1-), all adults showed corresponding BC discriminations (B1C1+/B1C2-, B2C2+/B2C1-) (Smeets, Barnes-Holmes, & Cullinan, 2000). In fact, Augustson, Dougher, and Markham (2000) reported that, after being trained on nine AB-C tasks (A1B1-C1, A1B3-C2, A1B2-C3; A2B3-C1, A2B2-C2, A2B1-C3; A3B2-C1, A3B1-C2, A3B3-C3), adults readily matched BC compounds with common A-elements (e.g., B1C1-B2C3). This account is consistent with the ideas put forward by Dougher and Markham (1994) and Stromer et al. (1993). These authors suggested that stimulus equivalence may be the result of a process of establishing discriminative control by separable compounds. The present analysis suggests that this process may also lead to equivalence-equivalence. Future research, therefore, should be directed at (a) identifying the process on which equivalence-equivalence is based: matching same discriminative functions or matching functionally same stimulus-stimulus relations, and (b) adapting the equivalence-equivalence measurements more closely to the relations measured by analogy tasks. Both these issues are currently under investigation.

References

ADAMS, B. J., FIELDS, L., & VERHAVE, T. (1993). Effects of test order on intersubject variability during equivalence class formation. *The Psychological Record*, 43, 133-152.

ALEXANDER, P. A., WILLSON, V. L., WHITE, C. S., FUQUA, J. D., CLARK, G. D., WILSON, A. F., & KULIKOWICH, J. M. (1989). Development of analogical reasoning in 4- and 5-year-old children. *Cognitive Development*, 4, 65-88.

AUGUSTSON, E. M., DOUGHER, M. J., & MARKHAM, M. R. (2000). Emergence of conditional stimulus relations and transfer of respondent eliciting functions among compound stimuli. *The Psychological Record*, 50, 745-770.

BAER, D. M., PETERSON, R. F., & SHERMAN, J. A. (1967). The development of imitation by reinforcing behavioral similarity to a model. *Journal of the Experimental Analysis of Behavior*, 10, 405-416.
BARNES, D. (1994). Stimulus equivalence and relational frame theory. *The Psychological Record*, 44, 91-124.
BARNES, D., HEGARTY, N., & SMEETS, P. M. (1997). Relating equivalence relations to equivalence relations: A relational framing model of complex human functioning. *The Analysis of Verbal Behavior*, 14, 57-83.
BARNES-HOLMES, Y., BARNES-HOLMES, D., ROCHE, B., & SMEETS, P. M. (2001). Exemplar training and a derived transformation of function in accordance with symmetry. *The Psychological Record*, 51, 287-308.
BOVET, D., & VAUCLAIR, D. (2001). Judgement of conceptual identity in monkeys. *Psychonomic Bulletin & Review*, 8, 470-475.
CARPENTIER, F., SMEETS, P. M., & BARNES-HOLMES, D. (2000). Matching compound samples with unitary comparisons: Derived stimulus relations in adults and children. *The Psychological Record*, 50, 671-685.
CARPENTIER, F., SMEETS, P. M., & BARNES-HOLMES, D. (2002). Matching functionally same relations: Implications for equivalence-equivalence as a model for analogical reasoning. *The Psychological Record*, 52, 351-370.
CULLINAN, V. A., BARNES-HOLMES, D., & SMEETS, P. M. (2001). A precursor to the relational evaluation procedure: Searching for the contextual cues that control equivalence responding. *Journal of the Experimental Analysis of Behavior*, 76, 339-349.
DARCHEVILLE, J. C., LEGRAND, P., & SMEETS, P. (1998). Symétrie et transitivité dans les stimulus composés discriminatifs. *Acta Comportamentalia*, 6, 129-142.
DOUGHER, M. J., & MARKHAM, M. R. (1994). Stimulus equivalence, functional equivalence, and the transfer of function. In S. C. Hayes, L. J. Hayes, M. Sato, & K. Ono (Eds.), *Behavior analysis of language and cognition* (pp. 71-90). Reno, NV: Context Press.
FIELDS, L., VARELAS, S., REEVE, K. F., BELANICH, J., WADHWA, P., DEROSSE, P., & ROSEN, D. (2000). Effects of prior conditional discrimination, symmetry, transitivity, and equivalence testing on the emergence of new equivalence classes. *The Psychological Record*, 50, 443-466.
GALLAGHER, J. M., & WRIGHT, R. J. (1979). Piaget and the study of analogy: Structural analysis of items. In J. Margery (Ed.), *Piaget and the helping professions* (Vol. 8, pp. 114-119). Los Angeles: University of Southern California.
GOSWAMI, U. (1991). Analogical reasoning: What develops? A review of research and theory. *Child Development*, 62, 1-22.
GOSWAMI, U., & BROWN, A. L. (1989). Melting chocolate and melting snowmen: Analogical reasoning and causal relations. *Cognition*, 35, 69-95.
GOSWAMI, U., & BROWN, A. L. (1990). Higher-order structure and relational reasoning: Contrasting analogical and thematic relations. *Cognition*, 36, 207-226.
HOLTH, P., & ARNTZEN, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record*, 48, 81-110.
LEVINSON, P. J., & CARPENTER, R. L. (1974). An analysis of analogical reasoning in children. *Child Development*, 45, 857-861.
LUNZER, E. A. (1965). Problems of formal reasoning in test situations. In P. H. Mussen (Ed.), European research in child development. *Monographs of the Society for Research in Child Development*, 30 (2, Serial No. 100).

PIAGET, J., MONTANGERO, J., & BILLETER, J. (1977). Les correlats. In J. Piaget (Ed.), *L'Abstraction reflechissante*. Paris: Presses Universitaires de France.

SAUNDERS, K. J., & WILLIAMS, D. C. (1998). Do parakeets exhibit derived stimulus control? Some thoughts on experimental control procedures. *Journal of the Experimental Analysis of Behavior*, 70, 321-324.

SAUNDERS, R. R., & GREEN, G. (1992). The nonequivalence of behavioral and mathematical equivalence. *Journal of the Experimental Analysis of Behavior*, 57, 227-241.

SCHUSTERMAN, R., & KASTAK, D. (1993). A California sea lion (*Zalophus californianus*) is capable of forming equivalence relations. *The Psychological Record*, 43, 823-839.

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

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

SIDMAN, M., & TAILBY, W. (1982). Conditional discrimination vs. matching-to-sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37, 5-22.

SMEETS, P. M., BARNES-HOLMES, D., & CULLINAN, V. (2000). Establishing equivalence classes with match-to-sample format and simultaneous-discrimination format conditional discrimination tasks. *The Psychological Record*, 50, 721-744.

STERNBERG, R. J., & RIFKIN, B. (1979). The development of analogical reasoning processes. *Journal of Experimental Child Psychology*, 27, 195-232.

STEWART, I., BARNES-HOLMES, D., ROCHE, B., & SMEETS, P. M. (2001). Generating derived relational networks via the abstraction of common physical properties: A possible model of analogical reasoning. *The Psychological Record*, 51, 381-408.

STOKES, T. F., & BAER, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, 10, 349-367.

STROMER, R., MCILVANE, W. J., & SERNA, R. W. (1993). Complex stimulus control and equivalence. *The Psychological Record*, 43, 585-598.

THOMPSON, R. K., & ODEN, D. L. (2000). Categorical perception and conceptual judgements by nonhuman primates: The paleological monkey and the analogical ape. *Cognitive Science*, 24, 363-396.