JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2011, **96,** 243–260 NUMBER 2 (SEPTEMBER)

TESTING FOR TRANSITIVE CLASS CONTAINMENT AS A FEATURE OF HIERARCHICAL CLASSIFICATION

Brian Slattery, Ian Stewart, and Denis O'Hora

NATIONAL UNIVERSITY OF IRELAND GALWAY

Three experiments investigated responding consistent with transitive class containment, a feature of hierarchical classification. Experiment 1 replicated key components of a preliminary attempt to model hierarchical classification (Griffee & Dougher, 2002) and tested for responding consistent with transitive class containment. Only 2 out of 5 participants showed the expected pattern. Experiment 2 tested whether repeated exposures to the Experiment 1 protocol would give rise to the expected pattern more reliably. None of 3 novel participants demonstrated the pattern. In Experiment 3, physically similar stimuli used in Experiments 1 and 2 were replaced across testing cycles by arbitrary stimuli. Transitive-class-containment-consistent responding was observed in all 3 novel participants. Implications, limitations and future research are discussed.

Key words: stimulus control, hierarchical classification, categorization, transitive class containment, humans

---

Classification refers to grouping stimuli into separate classes or categories on the basis of shared functions or physical characteristics (Astley, Peissig, & Wasserman, 2001). Hierarchical classification is a phenomenon in which classes themselves are categorized into higher order classes. An example of hierarchical classification is classifying ''Springer spaniel'' into the category ''Dog'', while classifying ''Dog'' itself into the category ''Mammal''.

Basic-level or within-class membership refers to classifying a stimulus as a member of a class without referring to a hierarchical taxonomy or the relationships between stimuli at different levels of such a taxonomy, for example, discriminating between two sets of stimuli on a perceptual basis (e.g., see Quinlan & Dyson, 2008). A considerable amount of behavioral research has been accrued on basic-level classification. This work has been conceptualized and modeled in terms of a number of behavioral processes including primary stimulus generalization (categories based on physical similarity) and functional and stimulus equivalence (categories not based on physical similarity; e.g., Galizio, Stewart, & Pilgrim, 2001, 2004; Lane, Clow, Innis, & Critchfield, 1998).

Address correspondence to Brian Slattery, School of Psychology, NUI Galway (e-mail: brianslattery2001@yahoo.com).

doi: 10.1901/jeab.2011.96-243

Most research on hierarchical (as opposed to basic) classification has been conducted by mainstream cognitive developmental researchers (e.g., Deneault & Ricard, 2006; Greene, 1989, 1994). The general consensus in this literature is that hierarchical classification develops over time and is characterized by three features: transitive class containment, asymmetry and property inheritance (e.g., Murphy, 2002; Winer, 1980). Transitive class containment refers to a pattern of responding whereby if someone is taught that an object (C) is a member of a class (B) and knows or is taught that class B is itself a member of class (A), then she should respond to C as a member of class A also. For example, if someone is taught that a Springer spaniel (C) is a dog (B), and that a dog is a mammal (A), then she should respond, without being taught, that a Springer spaniel (C) is a mammal (A).

Asymmetry, the second feature of hierarchical classification, refers to the phenomenon whereby class A contains class B, but class B does not contain class A (e.g., ''mammals'' contains ''dogs'', but ''dogs'' does not contain ''mammals''). The third feature is property inheritance. The properties of a superordinate class are shared with all members of subordinate classes, but all properties of the subordinate classes are not shared with members of a superordinate class. For example, all dogs are warm-blooded, but not all mammals have four legs. Cognitive developmental research typically uses qualitative inferential questions to assess the extent to which children of different ages begin to respond in accordance with different features of hierarchical classification (Deneault & Ricard, 2006; Greene, 1989; 1994; Winer, 1980). For example, a question probing transitive class containment might be, ''All Springer spaniels are dogs, and all dogs are mammals; are all Springer spaniels mammals?'' By analyzing responses to questions such as this, researchers have attempted to establish at what age and in what order the features of hierarchical responding emerge. Deneault and Ricard, for instance, used qualitative inference questions to assess transitive class containment and asymmetry and found that most 5-year-olds could respond correctly to probes for the former but not for the latter whereas most 9-year-olds responded correctly to both.

Cognitive research with adults has compared response latencies for different types of questions on hierarchically structured classes in order to test predictions concerning the structure of semantic memory. For example, according to the cognitive literature, each level of a hierarchy (e.g., Springer spaniel, dog, mammal) may be seen as a node and it has been found that response times differ depending on nodal distance. For instance, it has been found that people are faster to identify a phenomenon (e.g., Springer spaniel) as a member of a class such as ''Dog'' than as a member of a superordinate class such as ''Mammal'' (Murphy, 2002; Rips, 1975; Rips, Shoben, & Smith, 1973).

Though cognitive researchers have noted some key features of hierarchical classification and have gathered a variety of data on developmental changes and other variables that may be related to it, the theoretical explanations provided are inadequate from a behavior analytic perspective. Rather than focusing on identification of environment–behavior relations, they tend to employ hypothetical constructs (e.g., semantic memory) and thus are not readily translatable to applied settings (see Margolis & Laurence, 2000; Murphy, 2002; Palmer, 2002). As suggested above, there has been little or no behavior-analytic research on hierarchical classification, with the exception of Griffee and Dougher (2002), who conducted a preliminary investigation of hierarchical classification from a behavioral perspective.

Griffee and Dougher (2002) investigated the conditions necessary to establish responding in accordance with a natural language hierarchy. The latter includes stimuli that share common physical features and stimulus functions. An example is the taxonomy ''Springer spaniel *member of class* Dog; Dog *member of class* Mammal.'' People commonly respond to objects and events in accordance with hierarchical groupings such as this under the control of a combination of pertinent functions, physical characteristics and context. For example, I might respond differently to my Springer Spaniel and that of my neighbor based on particular physical features of each that I have learned to recognize. Other people might not be able to distinguish them so easily. However, they would readily discriminate my two dogs, a Springer Spaniel and an Alsatian, from each other since the latter differ physically to a much greater extent than two Springer Spaniels. This greater physical difference means that these two breeds tend to be responded to differently in more contexts than a pair of Springer Spaniels would be. Nevertheless, there are also many contexts in which a spaniel and an Alsatian are treated the same. For example, in assessing their general living requirements, I tend to respond to both as dogs, while in discussing particular behavioral similarities with humans, I might refer to them as mammals.

To empirically model features of responding in accordance with natural language hierarchical classes such as these, Griffee and Dougher (2002) established contextual control over responding to four similar (triangle) stimuli that differed along a physical continuum with respect to one feature (i.e., length of base). These stimuli represented objects or events that might be placed in a taxonomy such as that just discussed. In addition, color background was manipulated to represent a range of different contexts: superordinate, intermediate, and subordinate. Across all contexts, participants received points (exchangeable afterwards for money) as reinforcement for correct responding and received more points the more accurate their responding.

The green background served as a superordinate context. Participants were presented with four different-sized triangles as samples and seven nonsense syllables (which may be represented by the alphanumeric codes H1, H2a, H2b, H3ai, H3aii, H3bi, H3bii) as comparisons. They were trained to choose the same nonsense syllable (H1) from the comparison array in the presence of each of the four triangles. This modeled responding to a broad class of stimuli (e.g., responding to all members of an array of different animals as mammals). In the green background, the participants received six points (the maximum) for choosing H1 regardless of the sample and responses to the other six comparisons were not reinforced.

The red background served as the intermediate context. Participants were trained to choose two other nonsense syllables (H2a and H2b) from the same seven-syllable comparison array. Choices of H2a were reinforced in the presence of the two smaller-base triangles and choices of H2b were reinforced in the presence of the two larger-base triangles. Correct choices of H2a or H2b were reinforced with six points. However, choices of H1 after any of the samples were reinforced with three points. In this way, the researchers modeled the occurrence of more or less precise responding as it could occur in natural language. Correct choices of H2a or H2b in the red background are the most precise because they are analogous to responding to a stimulus using an intermediate class term (e.g., dog, cat) in a context in which that is appropriate (e.g., talking about what kind of pets I own). Choices of H1 in the red background are less precise and are analogous to using a superordinate class term (e.g., saying ‘‘mammal’’ when talking about my pets) in the intermediate context.

Finally, the yellow background served as a subordinate context. Participants were trained to choose a particular nonsense syllable from the four remaining nonsense syllables (i.e., H3ai, H3aii, H3bi and H3bii) in the presence of each of the four triangles respectively. This modeled responding to smaller (more specific) subclasses (e.g., responding differently to each of two different breeds of dog and two different breeds of cat). Participants received six points for choosing H3ai, H3aii, H3bi or H3bii in the presence of the appropriate sample. They received three points for choosing H2a or H2b in the presence of the smaller-base or larger-base triangle samples, respectively, and they received one point for choosing H1 in the presence of any of the samples.

The participants were then tested with novel triangle stimuli that differed from the originals in terms of base length, and they duly showed generalization of the response pattern. Next, in order to simulate a response function other than naming the stimuli, the participants were trained to press onscreen buttons under contextual control and generalization was again tested. Finally, to introduce arbitrary (language-like) relations into the model, derived relations between the nonsense syllables and buttons and the nonsense syllables and triangles were tested.

Thus, Griffee and Dougher (2002) provided an empirical demonstration of some key features of responding in accordance with hierarchical classes. In particular, the core phase of their study showed contextual control over the number of stimuli that participants responded to in a similar way. However, this represents only a preliminary model of particular features of hierarchical classification. Further investigation of possibly important features as discussed in cognitive literature, for example, is required. For instance, as noted above, hierarchical classification appears to involve the characteristics of transitive class containment, asymmetry and unidirectional property inheritance. If the performance of participants in Griffee and Dougher was functionally similar to the responding shown by participants in cognitive developmental research, for example, then participants trained as in Griffee and Dougher should also show behavior consistent with transitive class containment, asymmetry and unidirectional property inheritance.

Consider transitive class containment for instance. As noted earlier, transitive class containment refers to a response pattern whereby if someone is taught that an object (C) is a member of a class (B), then she will also respond that C is part of a second class (A) that subsumes class B. Transitive class containment seems an appropriate initial target of testing since empirical studies report that it is seen earlier in development than other features of hierarchical classification (e.g., Greene, 1989; 1994).

Griffee and Dougher (2002) did not directly investigate whether participants' responding was consistent with transitive class containment. However, some evidence was observed in the participants' errors (see pp. 442-443).

One way to test for transitive class containment in a natural-language context is to assess responding to hypothetical additional members of previously established classes. For example, participants could be introduced to a novel stimulus in one context (controlling classification at one level of the taxonomy) and then be tested for responding in a different context (controlling classification at a more general level). For instance, they might be told that an X is a type of dog and then asked whether an X is a mammal or a bird. A similar strategy might be used to assess if participants exposed to the Griffee and Dougher protocol would subsequently respond consistent with transitive class containment. Specifically, participants might be trained to respond in a particular way to a novel stimulus in one context (e.g., the subordinate yellow background context, which should control responding to stimuli in a relatively specific way) and then tested (i.e., in the absence of reinforcement) for responding to the same stimulus in a different context (e.g., the intermediate red background context, which should control responding to stimuli in a relatively more general way). Griffee and Dougher did test for responding to novel stimuli in different contexts employed in their protocol. However, their tests always used stimuli that were physically similar to the training stimuli and the observed performances likely were the result of stimulus generalization. In order to provide a more conclusive test of transitive class containment, stimuli that are physically distinct from the training stimuli must be employed.

EXPERIMENT 1

One key aim of this study was to assess whether participants trained using the Griffee and Dougher (2002) procedure would show transitive-class-containment-consistent responding. Phase 1 replicated the training structure employed by Griffee and Dougher, and consisted of a series of related conditional discriminations followed by a generalization test to determine whether the trained contextually-controlled response pattern would occur in the presence of novel stimuli. Phase 2 then tested whether responding was consistent with transitive class containment. Because the focus in this test was on responding in accordance with putative levels of a hierarchy, the nomenclature typically used to describe stimuli in hierarchical classes (i.e., subordinate class level, intermediate class level, superordinate class level) will sometimes be employed to refer to the different contexts.

In Phase 2, a group of four novel abstract stimuli (A1, A2, A3 and A4; referred to as Group 1) replaced the triangle stimuli as conditional stimuli in the subordinate background context. In the presence of each Group 1 stimulus, participants were trained to choose one of the four previously trained subordinate level nonsense syllables (i.e., H3ai, H3aii, H3bi, H3bii). Then, in Phase 2 testing, Group 1 stimuli were presented as conditional stimuli in the intermediate level background context to test whether participants would choose the appropriate intermediate level nonsense syllables (i.e., H2a or H2b). Such responding would be consistent with transitive class containment (see Figure 1).

Method

Participants

Five experimentally naïve females took part in Experiment 1. Their ages ranged from 22 to 28 years (M = 24.6 years; SD = 4.7 years). All participants signed a statement of informed consent before the experiment and were informed that there would be a 30 euro prize for the participant with the highest score at the end of the study.

Settings, Materials, and Apparatus

Each participant was tested individually in a cubicle that contained a desk, chair and personal computer (Fujitsu Siemens Scenic). Instructions, stimulus presentation and recording of responses were controlled by the computer, which was programmed in Visual Basic 2008. The experimental stimuli used in Phase 1 were similar to those used by Griffee and Dougher (2002). The first set of stimuli included eight triangles that differed in terms of the angle between sides A and B and the length of the base C (see Figure 2). The length of the sides A and B was always 2 inches (5.08 cm), and the angles between sides A and B ranged from 24° to 168° (T1 = 12°, T2 = 24°, T3 = 60°, T4 = 72°, T5= 108°, T6 = 120°, T7 = 156°, T8 = 168°). The triangles were black and were presented on a green, red, or

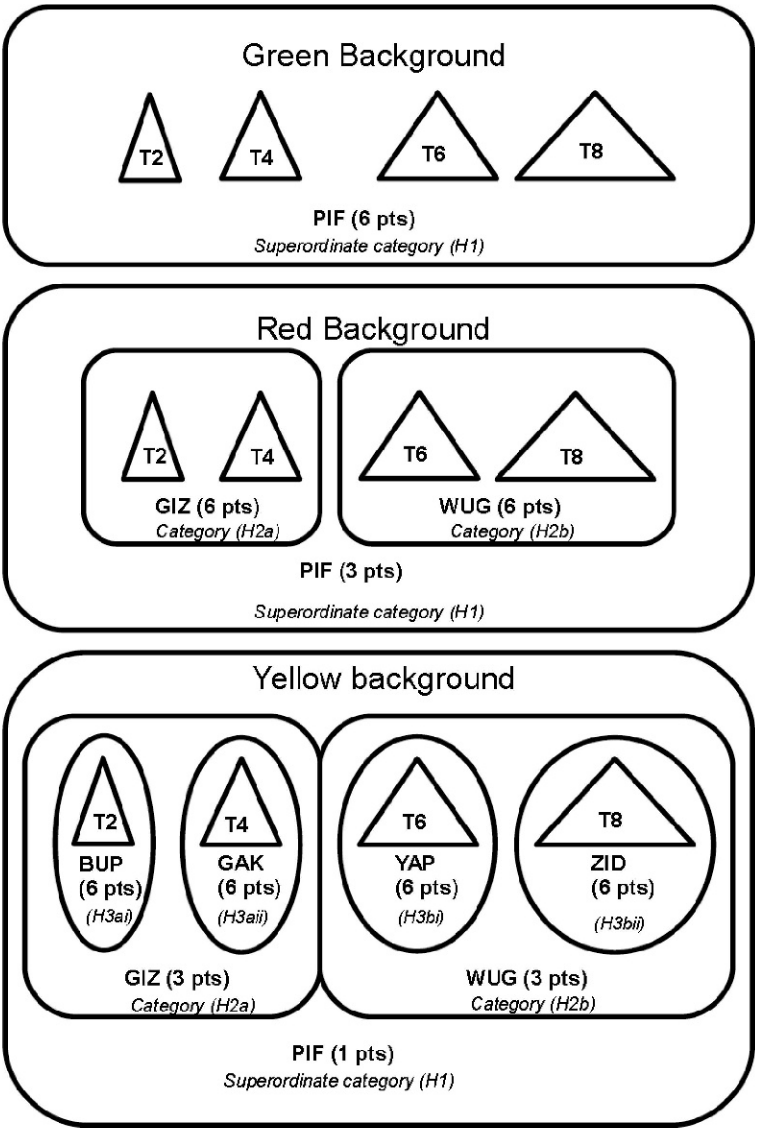

Fig. 1. Diagram illustrating the experimenter-designated hierarchical classification of stimuli. The appropriate triangle and nonsense syllable matches based on background color are shown, as well as the points available for correct responses during training.

yellow background. The seven nonsense syllables were PIF (H1), GIZ (H2a), WUG (H2b), BUP (H3ai), GAK (H3aii), YAP (H3bi) and ZID (H3bii). For Phase 2 (transitive class containment) training and testing, four novel abstract shapes were employed (Group 1: A1, A2, A3, A4; see Appendix).

Procedure

Phase 1: Conditional Discrimination Training. At the start of conditional discrimination training, participants were presented with the following instructions:

> On the computer screen you will see a background color, a triangle at the top of the screen and seven nonsense syllables at the bottom of the screen. You earn points by choosing the correct button, depending on what you see on the screen. It might take a while to learn which is the correct button, because the triangles are similar. But there will always be a ''best'' answer, which is worth

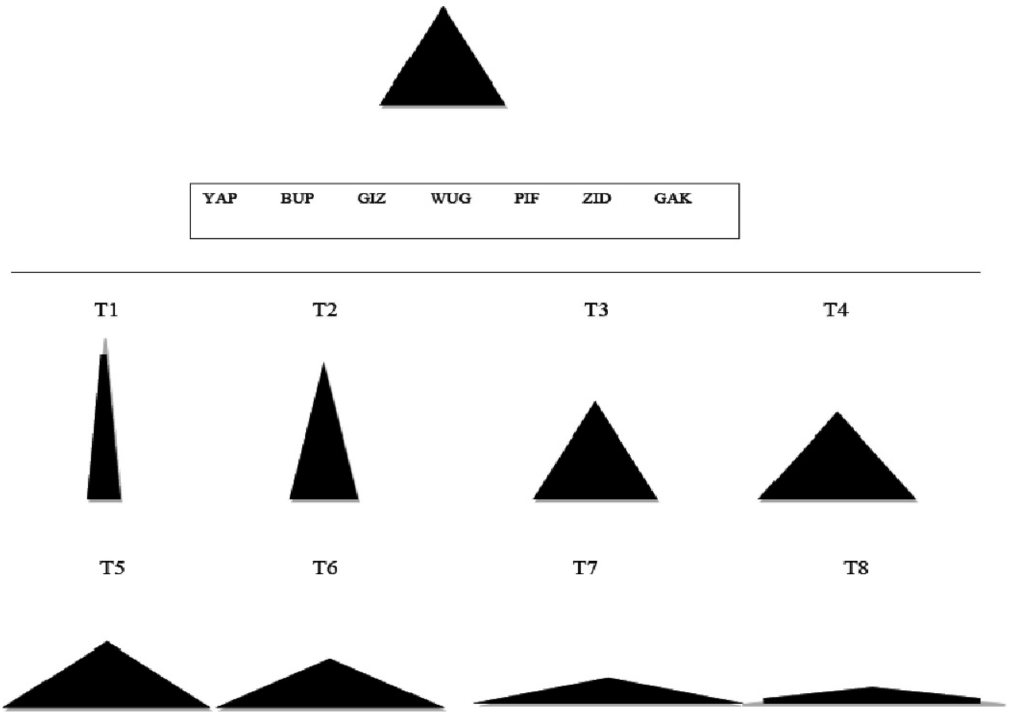

Fig. 2. Triangle stimuli used in the core protocol (bottom) and an example of a sample trial used during training in Phase 1 (top).

> 6 points. Sometimes there will also be a ''next best answer'' worth three points, and a ''next-next best'' answer, worth 1 point. The rest of the answers will be wrong. Your job is to earn as many points as possible! We will keep track of the points over the course of the semester and the person with the most points will win thirty euro.

The participants were taught to choose particular nonsense syllables in the presence of each of four different triangle samples (i.e., T2, T4, T6, T8). The correct choice was contingent both on the particular triangle sample presented and the particular screen background color (green, red or yellow). The four triangles were presented in quasirandom order. On each trial, a triangle was presented in the top center of the screen on one of the three background colors. The appearance of the sample was immediately followed by seven nonsense syllables presented across the bottom of the screen (see Figure 2, top). The right-to-left position of the seven nonsense syllables was counterbalanced across trials.

The participant used the mouse to select one of the seven nonsense syllables. Depending on the participant's response, the feedback presented was: ''6 points. This is the best answer''; ''3 points. But there is a better answer''; ''1 point. There are better answers''; or ''Incorrect''. Once the participant had made a response, the screen cleared, feedback was delivered and, after a 1-s ITI, the next trial began.

In the green background, the correct response (six points; see Figure 1) was H1 after all four samples. All other responses were incorrect. When the background color was red, participants earned six points for selecting H2a in the presence of T2 and T4 and H2b in the presence of T6 and T8. If the participant selected H1 in the red background, they earned three points. All other responses were incorrect. When the background color was yellow, participants earned six points for selecting H3ai in the presence of T2, H3aii in the presence of T4, H3bi in the presence of T6 and H3bii in the presence of T8. If the participant selected H2a in the presence of T2 and T4 or H2b in the presence of T6 and T8, they earned three points. If the participant selected H1 in the yellow background, they earned one point. All other responses were incorrect.

The current experiment used the same mastery criteria during conditional discrimination training and testing as Griffee and Dougher (2002). In addition, participants could also pass training blocks on the basis of a predetermined number of consecutively correct responses (cf., Gómez, López, Baños-Martin, Barnes-Holmes, and Barnes-Holmes, 2007).

Initial training trials were in the superordinate context and were presented with a green background. The training criterion in the green background was 95% correct responses after a minimum of 50 trials or 20 consecutively correct six-point responses. The maximum number of trials per exposure was 110. No three-point or one-point responses were possible in the green background. If the participant did not meet criterion, then the training phase was repeated up to a maximum of three times. If a participant failed to meet criterion after four attempts, he or she was debriefed and thanked for taking part and his/her participation was discontinued.

Training in the intermediate phase with the red background was conducted next. The red background training criteria were at least 90% six-point responses and 5% or less three-point responses after a minimum of 50 trials and a maximum of 178 trials or 30 consecutively correct six-point responses. No one-point responses were possible in the red background. If the participant did not achieve the criteria, intermediate-context training was repeated up to a maximum of three times. If the criteria still were not met, then participation was discontinued.

Following the separate training phases with the green and red backgrounds, there was a mixed training phase in which trials from the green and red background were alternated in a quasirandom order across trials. The same training criteria that applied in the red background were employed during this phase. If the participants did not achieve the criteria, the mixed-training phase was repeated up to a maximum of three times. If they did not complete this phase within three exposures their participation was discontinued.

Following the mixed training phase, participants were trained in the subordinate context with a yellow background. In this phase, the criteria were at least 88% six-point responses, 5% three-point responses, and 2% one-point responses for a minimum of 50 trials and a maximum of 200 trials, or 40 consecutively correct six-point responses. If the participants failed to achieve the criteria, the training was repeated up to a maximum of three times. If they failed to meet criterion within four exposures, their participation was discontinued.

In the final mixed-training phase, trials from the green, red, and yellow backgrounds were alternated in a quasirandom order across trials. The same training criteria that applied in the yellow background phase were used here. The minimum number of trials was 150 and the maximum was 320. If the participants did not achieve the criteria, training was repeated up to a maximum of three times. If they failed to meet the criteria within four exposures, their participation was discontinued.

Phase 1: Generalization Test. The generalization test probed for responding to novel triangles on the basis of physical similarity. Participants were first provided with the following instructions:

> This part of the experiment is slightly different than before. Even though correct answers still earn the same points as before, you will not receive feedback after your choice. You will also see some new triangles on the screen. Even though there will be some new triangles and you won't see the points you earn, there is always a best, 6-point answer and sometimes other answers are worth 3 points or 1 point. Remember, pay attention to the background color and try to earn as many points as possible. Best of luck.

Generalization testing consisted of a series of trials presented without feedback to determine whether the trained pattern of responding to the original four triangles (T2, T4, T6, T8) would transfer to the novel triangles (T1, T3, T5, T7). The four original triangles were included in the testing. Each of the eight stimuli was presented quasirandomly five times in each of the three background colours, for a total of 120 test trials. In the green background, the correct response after all eight triangles was H1. In the red background, the correct response after T1, T2, T3, and T4 was H2a while the correct response after T5, T6, T7, and T8 was H2b. In the yellow back-

Table 1
Conditional discrimination training data for participants in Experiment 1.

	Green			Red			Green + Red			Yellow			Green + Red + Yellow		
Ps	Trial no	%	Pass type	Trial no	%	Pass type	Trial no	%	Pass type	Trial no	%	Pass type	Trial no	%	Pass type
P1	54	56	Con	146	65	Con	56	82	Con	166	47	Con	226	95	%
P2	91	40	Con	140	77	Con	50	96	%	85	76	Con	151	93	%
P3	111	75	Con	50	98	%	50	100	%	50	96	%	150	100	%
P4	49	61	Con	87	88	Con	50	100	%	98	63	Con	150	98	%
P5	49	73	Con	134	73	Con	50	98	%	171	82	Con	151	97	%

Note. Ps = participant; Trial no = number of trials in that particular background color before participant met the criterion; % = percentage correct trials in that particular background color; Pass type = manner in which participant achieved the criterion, either Con (consecutively correct trials) or % (percentage criterion achieved).

ground, the correct response after: T1 and T2 was H3ai, T3 and T4 was H3aii, T5 and T6 was H3bi and T7 and T8 was H3bii. When the generalization test was complete, participants progressed to Phase 2.

Phase 2: Training and Testing for responding consistent with Transitive Class Containment. Participants were first presented with the following instructions.

> This part of the experiment is similar to one you have completed before. On the computer screen, you will see a background color, a random shape at the top of the screen and seven nonsense syllables at the bottom of the screen. You earn points by choosing the correct button, depending on what you see on the screen. Responses will be either correct ''6 points'' or incorrect ''0 points''. Your job as always is to earn as many points as possible! When you have read the instructions and you are happy that you understand them, press the spacebar to proceed to the actual experiment. Remember - try to get as many points as possible. Good luck!

Phase 2 training. Participants were trained on conditional discriminations in which the Phase-1 nonsense syllables were comparisons and novel abstract shapes (A1, A2, A3, A4) were samples. Training took place in the subordinate context (i.e., the yellow background). The correct response after: A1 was H3ai, A2 was H3aii, A3 was H3bi and A4 was H3bii. The seven nonsense syllables were presented across the bottom of the screen on each trial, and their positional order was counterbalanced across trials. To minimize feedback regarding the hierarchical structure, the tiered point system used in Phase 1 was not employed during this phase. All correct responses resulted in the delivery of six points, while other responses were followed only by feedback indicating that the response was incorrect. Once the participant had achieved the criterion of 90% accuracy after a minimum of 50 trials or 40 consecutively correct six-point responses, testing began. Participants were exposed to a maximum of 200 trials.

Phase 2 testing. Participants first saw the following instructions:

> This phase is a testing phase. The shapes you have just seen will now be presented on a different color background. You will not receive feedback. However, there is always a correct response worth '6' points while incorrect responses are worth '0' points. Remember - your job is to earn as many points as possible. Good luck!

This test involved the quasirandom sequential presentation of each of the four abstract shapes from training (A1, A2, A3, A4) four times each in the intermediate context (i.e., the red background). The seven nonsense syllables were presented across the bottom of the screen on each trial with their positional order quasirandomly varied across trials. After A1 and A2 samples, selections of H2a were consistent with transitive class containment, whereas after A3 and A4 samples, selections of H2b were consistent with transitive class containment.

Results and Discussion

Phase 1: Conditional Discrimination Training. The results of Phase 1 (conditional discrimination training with triangles) are presented in Table 1. All participants met the accuracy criteria on their first exposure to each training

Table 2

Generalization test data for all participants in Experiment 1.

	Triangle Stimuli							
Participant	1 (12°)	2 (24°)	3 (60°)	4 (72°)	5(108°)	6 (120°)	7 (156°)	8 (168°)
					Green			
P1	H1(5)	H1(5)	H1(5)	H1(5)	H1(5)	H1(5)	H1(5)	H1(5)
P2	H1(5)	H1(5)	H1(5)	H1 (5)	H1(5)	H1(5)	H1(5)	H1(5)
P3	H1(5)	H1(5)	H1(5)	H1 (5)	H1(5)	H1(5)	H1(5)	H1(5)
P4	H1(5)	H1(5)	H1(5)	H1 (5)	H1(5)	H1(5)	H1(5)	H1(5)
P5	H1(5)	H1(5)	H1(5)	H1 (5)	H1(5)	H1(5)	H1(5)	H1(5)
					Red			
P1	H2a(4) *H3aii(1)*	H2a(4) *H3aii(1)*	H2a(3) *H3ai(2)*	H2a(5)	H2b(5)	H2(5)	H2b(5)	H2b(5)
P2	H2a(5)	H2a(5)	H2a(5)	H2a(5)	H2b(3) *H2a(2)*	H2b(4) *H2a(1)*	H2b(5)	H2b(5)
P3	H2a(5)	H2a(5)	H2a(5)	H2a(5)	H2b(5)	H2b(5)	H2b(5)	H2b(5)
P4	H2a(5)	H2a(5)	H2a(5)	H2a(5)	H2b(5)	H2b(5)	H2b(5)	H2b(5)
P5	H2a(5)	H2a(5)	H2a(5)	H2a(5)	H2b(5)	H2b(5)	H2b(5)	H2b(5)
					Yellow			
P1	H3ai(5)	H3ai(5)	H3aii(5)	H3aii(5)	H3bi(5)	H3bi(5)	H3bi(5)	H3bii(3) *H3bi(2)*
P2	H3ai(5)	H3ai(2) *H1(1)* *H2b(2)*	H3aii(5)	H3aii(5)	H3bi(5)	H3bi(5)	H3bii(4) *H2b(1)*	H3bii(5)
P3	H3ai(5)	H3ai(5)	H3aii(5)	H3aii(5)	H3bi(5)	H3bi(5)	H3bii(1) *H3bi(4)*	H3bii(5)
P4	H3ai(5)	H3ai(5)	H3aii(5)	H3aii(5)	H3bi(5)	H3bi(5)	H3bii(5)	H3bii(5)
P5	H3ai(5)	H3ai(5)	H3aii(5)	H3aii(5)	H3bi(5)	H3bi(5)	H3bii(5)	H3bii(5)

Note. Alphanumeric codes; H1 = PIF; H2a = GIZ; H2b = WUG; H3ai = BUP; H3aii = GAK; H3bi = YAP; H3bii = ZID. The numbers inside the parentheses denote the number of times participants made that selection response. Incorrect responses are in italics

phase within each of the three contexts. In addition, 4 of 5 participants (P1, P2, P4, P5) met the criteria on the basis of number of consecutively correct trials with one exception: in the mixed Green–Red–Yellow background training, all participants met the percentage correct criteria. P3 met the criteria on the basis of percentage correct in all phases but the initial phase with the green background.

Phase 1: Generalization Test

The data from the generalization test are presented in Table 2. The participants readily showed generalization of the contextually controlled response pattern to the novel triangle stimuli. No errors were observed in the superordinate context (i.e., the green background). In the intermediate context (i.e., the red background), the correct response was H2a for T1 and T3, and H2b for T5 and T7. Incorrect responses were observed in the case of just 2 participants. Participants 1 and 2 showed four and three (respectively) incorrect responses (out of 120 trials) that were apparently based on inappropriate control by background color (i.e., making a response in the red background that would have been appropriate in the yellow). In the subordinate context (i.e., the yellow background), the correct responses were H3ai in the presence of T1, H3aii in the presence of T3, H3bi in the presence of T5, and H3bii in the presence of T7. Participants 1 and 3 made a number of responses which appeared to involve a failure to discriminate between nonsense syllables associated with that particular level of the hierarchy. For example, P3 selected H3bi instead of H3bii in the presence of T7. Participant 2 made a number of responses that appeared to involve a failure to discriminate background color, such as selecting H2b instead of H3bii in the presence of T7. Despite this, the dominant pattern was of high levels of appropriate generalization. Thus, the core findings from Griffee and Dougher (2002) were replicated.

Phase 2: Testing for Transitive-Class-Containment-Consistent Responding

Data from the Phase 2 probes are presented in Table 3 and Figure 3. During training,

Table 3

Percentage selection of stimuli by participants during Experiment 1 Phase 2 testing.

Ps + NS	H1	H2a	H2b	H3ai	H3aii	H3bi	H3bii
P1							
A1/A2	0	87.5	12.5	0	0	0	0
A3/A4	0	0	87.5	0	12.5	0	0
P2							
A1/A2	0	0	0	50	50	0	0
A3/A4	0	0	0	0	0	50	50
P3							
A1/A2	0	75	0	25	0	0	0
A3/A4	0	0	87.5	0	0	12.5	0
P4							
A1/A2	0	0	0	50	50	0	0
A3/A4	0	0	0	0	0	50	50
P5							
A1/A2	0	0	0	50	50	0	0
A3/A4	0	0	0	0	0	50	50

Note. The percentage of times each nonsense syllable (H1 = PIF, H2a = GIZ, H2b = WUG, H3ai = BUP, H3aii = GAK, H3bi = YAP, H3bii = ZID) was selected in the presence of each of the abstract sample stimuli in Group 1 (i.e., A1, A2, A3 and A4) in the red background (intermediate context) during the transitive-class-containment probe. The correct response after A1 and A2 was H2a, and the correct response after A3 and A4 was H2b. Ps = Participant; NS = novel sample stimulus.

participants were required to satisfy a mastery criterion of 90% correct in a block of 200 trials, or 40 consecutively correct responses. Three participants (P1, P2, and P5) required more than one exposure to the training task. An accuracy of 80% or above was deemed as passing the test for transitive class containment. On this basis, only 2 out of 5 participants (i.e., P1 and P3) demonstrated such responding.

In summary, all 5 participants in Experiment 1 successfully passed the conditional discrimination training and generalization test in Phase 1, which involved a replication of Griffee and Dougher (2002). To further test the Griffee and Dougher protocol, Phase 2 employed a test for responding consistent with transitive class containment. However, only 2 out of 5 participants passed this test. Since transitive class containment is seen in the cognitive literature as a key feature of hierarchical classification (e.g. Blewitt, 1994; Deneault & Ricard, 2006; Greene, 1994; Harnad, 1996), this suggests that at least 3 of these 5 individuals were not showing responding consistent with hierarchical classification as typically defined.

EXPERIMENT 2

Given the low level of responding consistent with transitive class containment seen in Experiment 1, the next aim of this study was to determine whether additional training might facilitate it. In Experiment 2, novel participants were provided repeated exposures to the conditional discrimination training and testing employed in Experiment 1 in an attempt to enhance the stimulus control of the different-colored background contexts.

Participants first received conditional discrimination training identical to that in Experiment 1. The subsequent generalization test that was conducted in Experiment 1 was omitted in Experiment 2 because it was deemed that (i) there was already sufficient evidence (both from Experiment 1 and Griffee & Dougher, 2002) that participants exposed to the conditional discrimination training would show generalization; (ii) this test was not needed as a test for transitive-class-containment-consistent responding; and (iii) omitting this test shortened the protocol to allow for readministration of training in the three contexts. After completing Phase 1 training, participants were immediately given a test for responding in accordance with transitive class containment. If a participant failed the test, then he or she was reexposed to the conditional discrimination training and test for transitive-class-containment-consistent responding with a new set of stimuli (i.e., Group

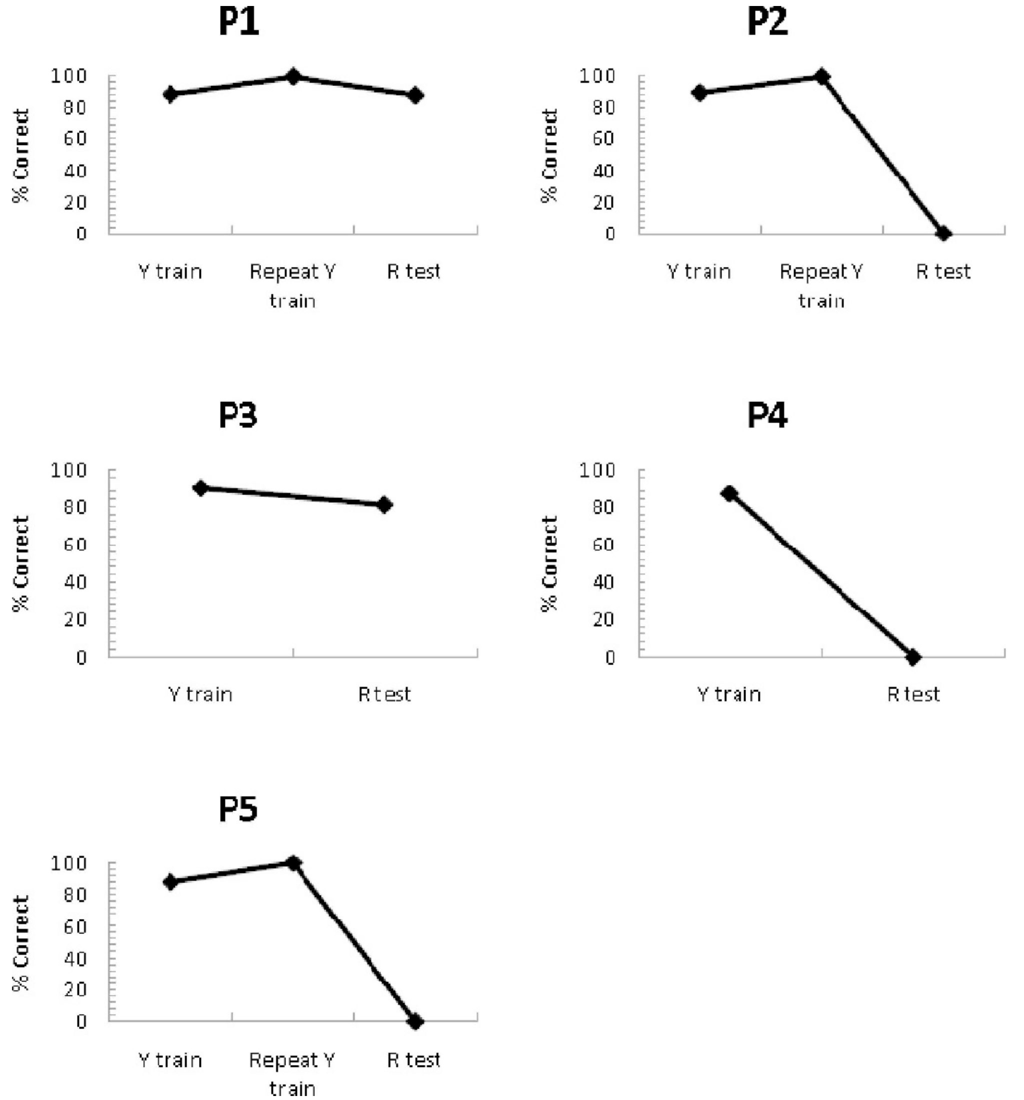

Fig. 3. Percentage of responses consistent with transitive class containment in training and testing in Phase 2 of Experiment 1. P1–P5 = Participant number. Y train = yellow background training (subordinate context); R test = red background testing (intermediate context).

2). This process continued until the participant met the criteria for transitive class containment or up to a maximum of five exposures (each with a different stimulus set).

Method

Participants

Three experimentally naïve participants (2 females and 1 male) took part in Experiment 2. Their ages ranged from 22-33 years (M = 26 years, SD = 5.9 years). The participants were informed that a 30 euro prize would be awarded to the person with the highest points score at the end of the study. All three signed a statement of informed consent before the experiment began.

Settings, Materials, and Apparatus

The setting, materials and apparatus used in Experiment 2 were almost exactly the same as in Experiment 1. Five groups of four novel abstract stimuli (Groups 1-5) were used (see Appendix).

Procedure

Phase 1 was identical to that in Experiment 1 except that no generalization test was conducted. Phase 2 began with exposure to training and testing for responding consistent with transitive class containment as in Experiment 1. However, if participants failed the test with the Group 1 novel abstract stimuli, they were reexposed to conditional discrimination training with the triangle samples and nonsense syllable comparisons in all three background contexts. Then, Phase 2 training and testing was repeated with a new set of abstract stimuli (Group 2) as samples. If they failed the Phase 2 test with these novel stimuli then they were reexposed to further conditional discrimination training in all three contexts, and Phase 2 training and testing was again repeated with additional novel stimuli (i.e., Group 3). This training-and-testing cycle was programmed to continue until either the participant had shown transitive-class-containment-consistent responding in the Phase 2 test or had been exposed to five cycles of testing without passing.

Results and Discussion

All participants passed conditional discrimination training and testing in Phase 1. Results for cycles of Phase 2 training and testing for the 3 participants are displayed in Figure 4. An accuracy level of 80% or above was required before a participant was deemed to have shown responding consistent with transitive class containment. Although all the participants continued to pass Phase 1 conditional discrimination training and successfully learned the required responses after each set of novel abstract shapes in the subordinate (i.e., yellow background) context during Phase 2 training, none of the 3 passed the test for responding consistent with transitive class containment. That is, in the intermediate (i.e., red background) context, no one participant made a single correct response. Instead, they continued to respond in accordance with the pattern learned during Phase 2 conditional discrimination training in the yellow background context.

In summary, all 3 participants in Experiment 2 failed to show transitive-class- containment-consistent responding despite repeated exposure to the conditional discrimination training employed in Griffee and Dougher (2002). Moreover, it is remarkable that none of the 3 participants showed any such responding whatsoever across multiple exposures to testing. One possible contributing factor might be the absence of the test for generalization. It is possible that having to choose novel triangle stimuli may make abstraction of the ''transitive class containment'' pattern more likely. Of course, this was not a part of the original Griffee and Dougher training protocol but was instead a test of that protocol. Furthermore, the affects of exposure to this test are at best limited since only 2 participants (P1 and P3) out of 5 demonstrated transitive class containment consistent responding even when it was employed in Experiment 1.

One reason that repeated exposure to the conditional discrimination training might not have produced transitive-class-containment-consistent responding is that exposure to this protocol may simply have reinforced responding under the control of physical (nonarbitrary) features of the stimuli employed in the training, and this source of control may have dominated over the more abstract (i.e., not based on physical characteristics of particular stimuli) pattern of transitive-class-containment-consistent responding required to pass testing. If it is the case that control based on physical features can dominate over a more generalized pattern of hierarchical classification, then this would mitigate against the possibility that the generalization test could have increased transitive class containment responding in Experiment 1, since generalization is based on physical similarity. If true, then the observed difference between levels of class-containment-transitive-consistent responding in Experiments 1 and 2 may have been a result of chance or else of some other unknown factor. One way to reduce control by physical features of the stimuli is to use abstract or arbitrary stimuli in the context of multiple exposures to the protocol. This approach was implemented in Experiment 3.

EXPERIMENT 3

Research has suggested that exemplar training aimed at establishing or tapping into a repertoire of generalized relational responding (including, for example, class hierarchical relations) should employ arbitrary stimuli

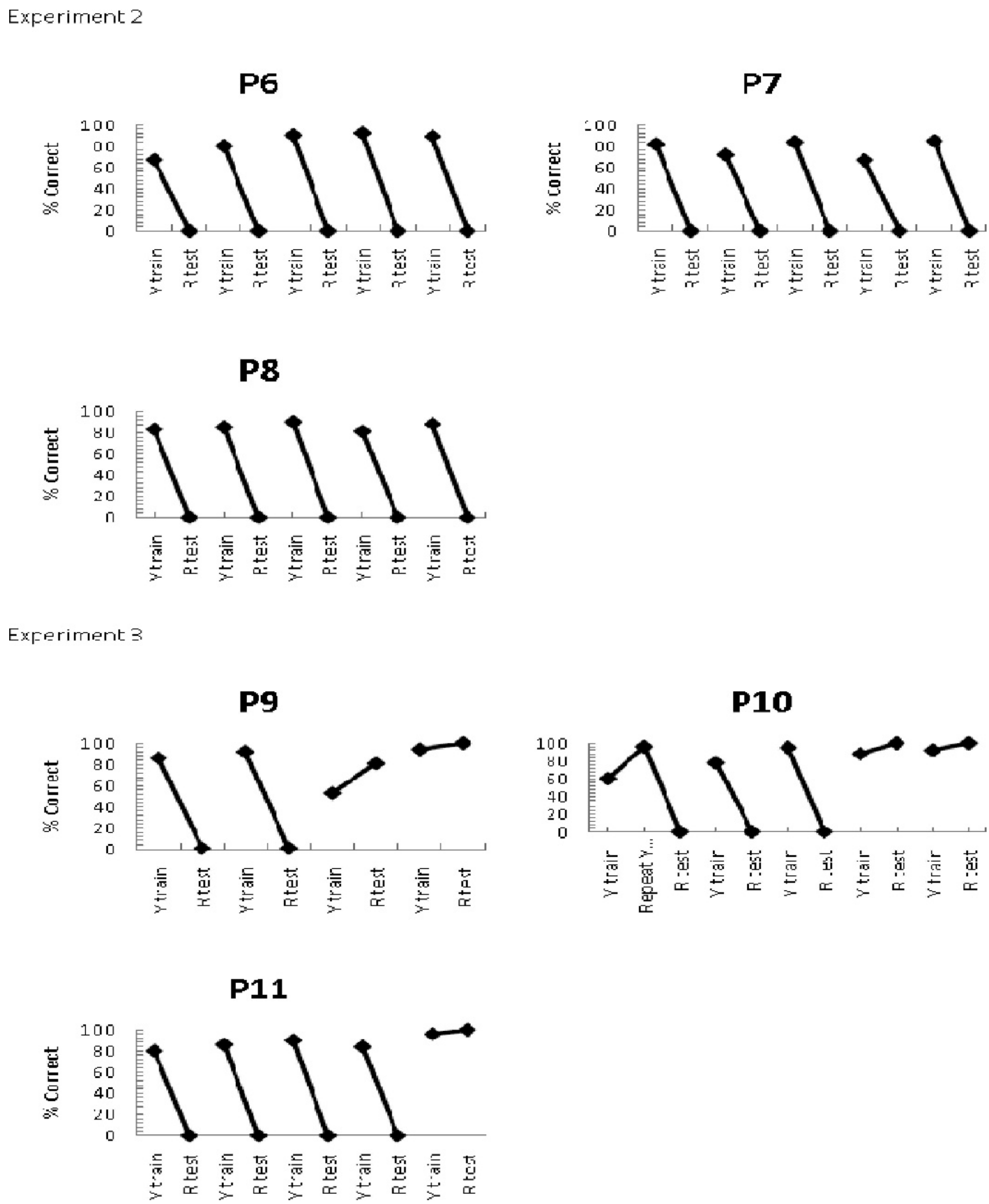

Fig. 4. Percentage of responses consistent with transitive class containment across exposures to novel exemplars in training and testing in Phase 2 of Experiments 2 and 3. P6 - P11 = Participant number. Y train = yellow background training (subordinate context); R test = red background testing (intermediate context).

which differ across exemplars, in order to mitigate the possibility of control by physical features of the stimuli (e.g., Barnes-Holmes, Barnes-Holmes, Smeets, Strand & Friman, 2004; Gómez et al., 2007). In Experiment 3, abstract stimuli were gradually introduced into the conditional discrimination training protocol. The introduction of abstract stimuli was gradual because previous work has suggested the importance of nonarbitrary control at least in the initial stages of training a relational pattern (e.g., Berens & Hayes, 2007).

Method

Participants

Three experimentally naïve participants (2 females and 1 male) took part in Experiment 3. They ranged in age from 24–30 years (M = 26.3 years, SD = 3.2 years). Participants signed a statement of informed consent before the experiment began and were debriefed once it was completed. Participants were told that 30 euro would be awarded to whoever had highest points at the end of the study.

Settings, Materials, and Apparatus

The setting, materials and apparatus employed in Experiment 3 were the same as in Experiment 2 except for the inclusion of a number of novel abstract stimuli in the Phase 1 conditional discrimination training phases (Group 6: N1, N2, N3, N4; see Appendix).

Procedure

Training and testing was similar to that for Experiment 2. The key difference was that the original four triangle-sample stimuli were systematically and progressively replaced over successive training and testing cycles by novel abstract stimuli in a kind of graduated training or fading procedure. Thus, in the first cycle of training, all the original stimuli (i.e., T2, T4, T6, T8) were employed; in cycle 2, stimulus T2 was replaced by an arbitrary stimulus (N1) so that the array of stimuli employed was N1, T4, T6 and T8; in cycle 3, the array was N1, T4, N3, T8; for cycle 4 it was N1, N2, N3 and T8; and for cycle 5 it was N1, N2, N3, and N4.

Results and Discussion

Results for cycles of Phase 2 training and testing for the 3 participants in Experiment 3 are displayed in Figure 4. An accuracy level of 80% or above was required before a participant was deemed to have passed testing for responding consistent with transitive class containment. As may be seen in Figure 4, all 3 participants passed the conditional discrimination training and then learned to choose the appropriate nonsense syllables in the presence of particular arbitrary shapes in the subordinate (i.e., yellow background) context in the training portion of Phase 2. However, all the participants failed the subsequent test for transitive-class-containment-consistent responding on their first exposure. Furthermore, they all continued to fail the test for transitive class containment in further Phase 2 tests as abstract sample stimuli began to replace the original triangle samples. However, eventually, after a number of exposures, each of the 3 participants passed this test.

In terms of individual performance, after her third cycle of training, P9 showed a substantial increase in transitive-class-containment-consistent responding (81% correct). By the fourth cycle, correct responding had risen to 100%. P10 also showed emergence of criterion responding (100%) by cycle 4, while P11 passed the test (100%) in cycle 5.

In summary, the procedure used in Experiment 3 successfully established responding consistent with transitive class containment in 3 experimentally naïve participants. This result was achieved using a procedure in which the participants were exposed to multiple cycles of conditional discrimination training similar to that used by Griffee and Dougher (2002), but with the critical added feature of the graduated inclusion of abstract stimuli in the array of training stimuli.

GENERAL DISCUSSION

In hierarchical classification, context determines the size of the groups of stimuli that are treated as functionally the same such that, in one context, a group of stimuli may all be responded to similarly (e.g., referring to a group of dogs and cats as mammals), whereas in another context, subgroupings of the stimuli in the overarching class may be responded to differently (e.g., referring to dogs and cats as distinct animals). Transitive class containment, which many theorists have agreed is a defining characteristic of hierarchical classification, refers to a response pattern whereby if someone is taught that an object C (e.g., a spaniel) is a member of a class B (e.g., dogs) and is taught that class B is itself a member of class A (i.e., that dogs are mammals), then, in the absence of training, the individual responds to C as a member of class A (i.e., that spaniels are mammals). The current research expanded on Griffee and Dougher's (2002) preliminary behavior-analytic exploration of features of hierarchical classification by testing whether, after exposure to that procedure, participants would pass a test for responding in accordance with transitive class containment.

In Experiment 1, after being exposed to the partial replication of the original Griffee and Dougher (2002) training protocol (Phase 1), only 2 out of 5 participants passed the Phase 2 test of transitive class containment. In Experiments 2 and 3, the protocol was altered in an attempt to facilitate responding consistent with transitive class containment. In Experiment 2, 3 novel participants were exposed to multiple rounds of training using repeated exposures to conditional discrimination training interspersed with testing using multiple sets of novel abstract stimuli. This repeated training, however, failed to establish transitive class containment. In Experiment 3, the original nonarbitrary stimuli from the initial conditional discrimination training were progressively replaced with novel arbitrary stimuli. This adaptation appeared to be successful in establishing reliable patterns of responding consistent with transitive class containment in 3 novel participants.

It might be argued that this protocol has demonstrated not transitive class containment but instead merely generalized color-background-controlled second-order conditional discriminative responding. Such an argument assumes that transitive class containment is something more than the latter. Nevertheless, even if one were to accept that participants' behavior was under functionally simpler control than transitive class containment, the current study still represents a useful advance. Hierarchical classification and its features including transitive class containment are a complex pattern of behaviors that are learned over many years. We know that, before children learn hierarchical classification, they first learn relatively simpler, more constrained versions of this behavior. Through further exposure to the socio-verbal environment, this behavior generalizes and comes under more complex stimulus control. Thus, whether one takes successful Phase 2 test performance to constitute transitive class containment per se or merely a precursor to it, the current study has allowed us to begin to identify the learning processes potentially involved in hierarchical classification. In so doing, it will inform the development of procedures to establish classification repertoires in young children and populations in which this repertoire is deficient.

Additional research is required to analyze in more detail the variables responsible for the increased probability of responding consistent with transitive class containment observed in Experiment 3. For instance, one key variable may have been the use of arbitrary stimuli. This was a deliberate difference from the Griffee and Dougher (2002) protocol, which focused primarily on the use of physically or nonarbitrarily related stimuli. It is true that many examples of hierarchical categories involve physical relations between the stimuli involved. For example, scientific classifications of animals typically involve increasing similarity between members of the same class the further down the hierarchy one goes. In addition, it is certainly the case that the development of a repertoire of hierarchical classification requires training with physically similar objects. With regard to the current research, additional work is needed to examine the extent to which training based on physical relations alone is sufficient to generate a repertoire of transitive class containment and hierarchical classification. Nevertheless, the hallmark of a fully developed repertoire of hierarchical classification is the ability to generate and to understand novel classes that are not dependent on physical relations between stimuli. A typical adult can hypothesize and name a number of new animal species and subspecies and provide an illustration of hierarchical relationships among these hypothetical groups without any mention of their actual physical characteristics and yet a listener trained in hierarchical classification might still be able to respond in accordance with new relations between those hypothetical groups based on the information provided. Such patterns of speaker and listener behavior constitute examples of responding in accordance with abstract hierarchical relations. The higher probability of transitive-class-containment- consistent responding observed in Experiment 3 may have been facilitated by the inclusion of abstract stimuli.

Stimulus classes in which the stimuli are not physically related may be useful in the exploration of hierarchical classification and its core features including transitive class containment, asymmetry and property inheritance. One way to proceed, for example, might be to train and test equivalence classes in one context and then train and test further equivalence classes under contextual control as subdivisions of the initial classes. It would be interesting to examine whether this approach to the creation of hierarchical classification might yield transitive-class-containment-consistent responding more readily than the above procedure, for example. In addition, this proposal to use contextual control to establish equivalence classes within equivalence classes might allow for useful exploration of aspects of stimulus equivalence as well as of hierarchical classification.

In addition to the use of arbitrary stimuli, the training protocol used in Experiment 3 involved the graduated (i.e., over the course of training cycles) fading out of the original nonarbitrarily (physically) related triangle stimuli and their replacement with abstract stimuli. However, further research is needed to investigate the extent to which the introduction of arbitrary stimuli or the use of a graduated procedure or a combination of both might have resulted in the success of the protocol used in Experiment 3. For instance, the efficacy of the procedure used herein might be compared with that of another in which exclusively arbitrary stimuli were used from the start of training.

Transitive class containment is only one of a number of characteristics that theorists have argued are needed for hierarchical classification. The other two features are asymmetry and unidirectional property inheritance. Procedures for testing these properties from a behavior-analytic perspective have yet to be developed. Both asymmetry and property inheritance appear to be more complex patterns of behavior than transitive class containment. Deneault and Ricard (2006) found that most 5-year-olds could respond correctly to probes for transitive class containment but not for asymmetry whereas most 9-year-olds responded correctly to both. With regard to property inheritance, there has been relatively less work conducted on this than on the other features of hierarchical classification, but it seems functionally more similar to asymmetry than to transitive class containment in at least one important respect. Though asymmetry and property inheritance are formally different in a number of ways, from a behavioral perspective, they are functionally similar in that both might be captured by modeling two different forms of contextual control—''member of'' and ''category of'', which determine different responses (e.g., in the presence of ''member of'' and ''dog'', choose ''spaniel'' not ''animal'').

One way to model hierarchical classification using ''member of'' and ''category of'' contextual control to test for asymmetry and property inheritance would be to train *hierarchical relational framing*. Hayes, Barnes-Holmes and Roche (2001) suggested that hierarchical classification is based on the training of generalized contextually controlled patterns of relational responding that originate in nonarbitrary relations of containment. For example, a child learns in one context to tact things as being physically inside other things (e.g., ''my hand is in my glove'') and in another context, to tact things as containing other things (e.g., ''the house contains the doll''). This repertoire then comes under contextual control (i.e., of cues such as the words ''in'' and ''contains'') and generalizes to more abstract behaviors such as classification in which these two repertoires (i.e., being contained [membership], versus containing [being a category or class]) are brought to bear. Relational Frame Theory (RFT) suggests that one way to investigate both asymmetry and property inheritance (and train them if absent) in the laboratory is to establish arbitrary stimuli as cues using nonarbitrary training (i.e., involving objects or pictures of objects being contained within or containing other objects) and then using those cues to gauge the relationship (e.g., ''member of?'', ''category of?'') between previously unseen novel stimuli, some of which might function as ''members'' and others as ''classes''.

It should also be noted that several participants in Experiment 1 showed responding consistent with transitive class containment even without the additional training provided in Experiment 3, and thus it is possible that relatively minimal aspects of the training and testing procedures are sufficient to bring a repertoire of hierarchical classification to bear for some participants. A number of features of the protocol might have been sufficient to do this, including contact with a series of closely physically related objects across training and testing trials, contact with differing levels of reinforcement within the same trial type for different stimulus choices, and contact with differing levels of reinforcement across different (color background) contexts. Those for whom such features more strongly cued transitive class containment responding may have been more likely to show this repertoire in a test with abstract stimuli, whereas those for whom it did not or for whom it cued transitive class containment only in the context of physically related stimuli were less likely to do so.

Given that adult human volunteers participated in the current study, it seems likely that the procedures reported here tapped into preexisting repertoires of hierarchical categorization rather than establishing such a repertoire a priori. Nevertheless, by bringing the behavior of these participants under the control of relevant protocols in the laboratory, we can gain a better understanding of the processes that appear to be involved in this apparently complex repertoire. As we develop more successful procedures for testing and training hierarchical classification and associated skills in this context, we then become able to use such procedures with populations, including young children or those with developmental disabilities, whose repertoire of complex behavior is much less advanced.

REFERENCES

Astley, S. L., Peissig, J. J., & Wasserman, E. A. (2001). Superordinate categorization via learned stimulus equivalence: Quantity of reinforcement, hedonic value, and the nature of the mediator. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*, 252–268.

Barnes-Holmes, Y., Barnes-Holmes, D., Smeets, P. M., Strand, P., & Friman, P. (2004). Establishing relational responding in accordance with more-than and less-than as generalized operant behavior in young children. *International Journal of Psychology and Psychological Therapy*, *4*, 531–558.

Berens, N. B., & Hayes, S. C. (2007). Arbitrarily applicable comparative relations: Experimental evidence for a relational operant. *Journal of Applied Behavior Analysis*, *40*, 45–71.

Blewitt, P. (1994). Understanding categorical hierarchies: The earliest levels of skill. *Child Development*, *65*, 1279–1298.

Deneault, J., & Ricard, M. (2006). The assessment of children's understanding of containment relations: Transitivity, asymmetry, and quantification. *Journal of Cognition and Development*, *7*, 551–570.

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

Galizio, M., Stewart, K. L., & Pilgrim, C. (2004). Typicality effects in contingency shaped generalized equivalence classes. *Journal of the Experimental Analysis of Behavior*, *82*, 253–273.

Gómez, S., López, F., Martín, C. B., Barnes-Holmes, Y., & Barnes-Holmes, B. (2007). Exemplar training and a derived transformation of functions in accordance with symmetry and equivalence. *The Psychological Record*, *57*, 273–294.

Greene, T. R. (1989). Children's understanding of class inclusion hierarchies: The relationship between external representation and task performance. *Journal of Experimental Child Psychology*, *48*, 62–89.

Greene, T. R. (1994). What kindergartners know about class containment hierarchies. *Journal of Experimental Child Psychology*, *57*, 72–88.

Griffee, K., & Dougher, M. J. (2002). Contextual control of stimulus generalization and stimulus equivalence in hierarchical categorization. *Journal of the Experimental Analysis of Behavior*, *78*, 433–447.

Harnad, S. (1996). Experimental analysis of naming behavior cannot explain naming capacity. *Journal of the Experimental Analysis of Behavior*, *65*, 262–264.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational Frame Theory: A post-Skinnerian account of human language and cognition.* New York: Kluwer Academic/ Plenum Publishers.

Lane, S. D., Clow, J. K., Innis, A., & Critchfield, T. S. (1998). Generalization of cross- modal stimulus equivalence classes: Operant processes as components in human category formation. *Journal of the Experimental Analysis of Behavior*, *70*, 267–279.

Margolis, E., & Laurence, S. (2000). *Concepts: core readings.* Cambridge, MA: The MIT Press.

Murphy, G. (2002). *The big book of concepts.* Cambridge, MA: The MIT Press.

Palmer, D. C. (2002). Psychological essentialism: A review of E. Margolis and S. Laurence (EDS.). *Journal of the Experimental Analysis of Behavior*, *78*, 597–607.

Quinlan, P., & Dyson, B. (2008). *Cognitive psychology.* Edinburgh: Pearson Education Limited.

Rips, L. J. (1975). Inductive judgements about natural categories. *Journal of Verbal Learning and Verbal Behavior*, *14*, 665–681.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*, 1–20.

Winer, G. A. (1980). Class inclusion reasoning in children: A review of the empirical literature. *Child Development*, *51*, 309–328.

Received: January 25, 2011
Final Acceptance: June 20, 2011

APPENDIX

Novel abstract stimuli used in the three experiments.