2013

Effects of CARO on Stimulus Equivalence: A Systematic Replication

Abdulrazaq Imam
John Carroll University, aimam@jcu.edu

J. V. Blanche

Effects of a CARO on Stimulus Equivalence Formation: A Systematic Replication

Abdulrazaq A. Imam and Justin V. Blanche
John Carroll University

In two experiments, we examined the disruptive effects of a "can't answer" response option (CARO) on equivalence formation. The first experiment was a systematic replication of Duarte, Eikeseth, Rosales-Ruiz, and Baer (1998), in which participants in a CARO group and a No-CARO group performed conditional discrimination tasks with abstract stimuli using a paper-and-pencil format for training and testing of equivalence relations. The presence of the CARO led to the nonemergence of equivalence classes. In the second experiment, participants performed conditional discrimination tasks using standard matching-to-sample training and testing procedures on a computer with CARO available only during testing. Equivalence yields were also low, with participants using CARO more on transitive and equivalence trials than on symmetry trials. The results support previous reports of equivalence disruption by nonresponse options such as CARO and suggest directions for further research.

Key words: stimulus equivalence, matching-to-sample, nonresponse options, CARO, touch screen, humans

Stimulus equivalence research tends to focus on factors that influence the development of equivalence classes. Usually of less concern, however, are the mechanisms that hinder the emergence of derived relations and the formation of equivalence classes. One variable that has been identified in the literature is giving participants the option of not responding to available class-based comparisons on some trials. Such nonclass-based response options in stimulus equivalence studies typically have included the presence of a comparison stimulus in addition to class-based comparisons that have been variously labeled "none" (e.g., Innis, Lane, Miller, & Critchfield 1998), "neither" (e.g., Fields, Adams, Brown, & Verhave 1993; Reeve & Fields, 2001), "can't answer" (henceforth, CARO; e.g., Duarte, Eikeseth, Rosales-Ruiz, & Baer, 1998), or "pass" (Saunders & Sherman, 1986). Although each of these terms signifies that the participant may choose not to select from among other given comparisons, they carry different implications, depending on the context. For example, the selection of a "none" and "neither" response can indicate that the appropriate answer exists but is not present among the comparisons. On the other hand, "pass" may indicate that the participant simply does not wish to select from the comparisons without

We thank Molly Gemp for her help in running some sessions in the second experiment and Lanny Fields, Richard Saunders, and Erik Arntzen for their helpful comments on an earlier version of the manuscript.

Justin V. Blanche is now at Cleveland State University. Correspondence concerning this article should be addressed to A. A. Imam, Department of Psychology, John Carroll University, 1 John Carroll Blvd., University Height, OH 44118. E-mail: aimam@jcu.edu

necessarily implying that an appropriate comparison is not present. Regardless of the interpretation or context, however, the nonresponse or default-response options appear to allow participants to respond in a manner that precludes establishing equivalence relations. Perhaps this happens as a result of the nonresponse option overshadowing the selection of a class-based comparison, even though such a selection would have occurred in its absence (see, e.g., Reeve & Fields, 2001). Reeve and Fields demonstrated just such an effect with dimensionally defined stimulus classes; dimensional variants that did not function as members of a perceptual class when an opt-out response was present did function as such when the opt-out response was not available.

Duarte et al. (1998) studied how presenting the CARO would affect the development of equivalence for a pair of three member classes in a match-to-sample (MTS) procedure. Four groups completed MTS tasks presented in a paper-and-pencil (PAP) format, using alphabets and numbers (A-1-X and B-2-Y) as stimuli. They employed a peculiar training procedure whereby participants received only instructions specifying rules on matching baseline relations as a template to which participants may refer at any point in baseline or testing, without any additional feedback. The groups varied by the restrictiveness of the instructions they received and by the presence of the option not to respond on certain trials or in testing. Their findings suggested that the CARO greatly impaired equivalence formation. Participants who had the CARO consistently failed on emergent relations regardless of how restrictive the instructions were. Eikeseth, Rosales-Ruiz, Duarte, and Baer (1997) reported a similar effect using an implied option not to respond by allowing participants to leave answers blank, again, using a paper-and-pencil format. They found that despite the lack of an explicit default-response option among the comparisons, 14% of participants did not respond to at least 15 of the 16 equivalence trials.

Prior to these studies, Fields et al. (1993) reported that when presented with a “neither” option during generalization tests in their second experiment, half the participants selected the “neither” option with two other class-based comparisons present. Innis et al. (1998) analyzed equivalence disruption with a default response option of “none.” In some of their groups, participants were trained to select “none” during catch trials. On these trials, none of the comparisons was correct, and the selection of the default response option resulted in positive feedback. Groups varied on the amount (none, 0%, 25%, and 50%) of catch-trial training participants received. They found that participants with greater amounts of catch-trial training more readily passed emergent relation tests that included a “none” response option than participants with less or no catch-trial training. Both studies employed computer-based standard equivalence procedures, presenting conditional discrimination tasks on a trial-by-trial basis.

A unique feature of the Duarte et al. (1998) study that is not in common use in standard equivalence research was that participants had access to the printed keys (the rules template mentioned above) for correct responses on baseline relations during testing. Coupled with the universal familiarity of the stimuli (letters and numbers: A-1-X and B-2-Y) used, the simultaneous availability of the notes, directions, and previously completed trials with each new trial are potential problems in that study because they were not only presented as illustrative instructions but also remained available for perusal for the entire session. In addition, because of the use of familiar stimuli, participants in that study essentially learned the two three-member equivalence classes by relying on their personal histories with the stimuli that were to become members of the equivalence classes. These various elements of their study afforded the participants in that study unique advantages for equivalence formation. The fact that the CARO participants were unable to form equivalence signifies, therefore, an important function of a default or no-response option in establishing equivalence classes, and thus warrants further empirical evaluation.

The present study sought to replicate the Duarte et al. (1998) study systematically by increasing the number of classes to three instead of two and thereby increasing the number of class-based comparisons from two to three, providing explicit feedback during training, and using abstract stimuli instead of familiar letters and numbers with their PAP format

(Experiment 1) as well as with standard computer-based MTS task using more extensive training in baseline, and making CARO available only during testing (Experiment 2). In Experiment 1, two groups participated as in the Duarte et al. study but worked with individual note cards depicting relevant training and testing relations instead of a booklet. One group had CARO available but the other did not during training and testing as arranged by Duarte et al. In Experiment 2, all participants had CARO available, with the criterion for training increased from four consecutive trials correct, used by Duarte el al. and in the first experiment, to 12 consecutive trials correct on individual training blocks. Based on prior studies cited previously, we expected the availability of CARO to diminish the likelihood of equivalence-class formation in both experiments but less so in Experiment 2 due to the additional training trials provided. Additionally, given Duarte et al.'s findings, we expected CARO availability to differentially engender poorer performance on transitive and equivalence relations than on symmetrical relations.

Experiment 1

The purpose of the first experiment was to replicate systematically the Duarte et al. (1998) study but modify their procedures in some important ways. We increased the number of classes from two to three to examine how adding another class-based comparison may have affected responses. Another important difference in this study is the inclusion of explicit standard equivalence training. In their original design, the baseline relations were displayed in an answer key that was available during the entire session, and participants were tested on baseline relations simultaneously with emergent ones in a single test booklet. There was no stated measure of mastery of baseline relations in the Duarte et al. procedure. Participants in Experiment 1 demonstrated mastery of the baseline relations before testing for the emergence of the derived relations. Consistent with Duarte et al.'s procedure, a group design was adopted for Experiment 1, and training of the baseline relations was sequential, following a linear-series (LS) training structure. All stimuli were abstract in order to avoid potential consequences of personal histories.

Method

Participants. Eighteen college students of at least 18 years of age from John Carroll University participated. Sixteen of them were enrolled in an introductory psychology course and received course credit; the remaining two volunteered for the study and received no credits for their participation. Participants had no prior experience in stimulus equivalence research.

Materials. Figure 1 presents the stimuli used in Experiment 1, obtained from the Microsoft Office character databank. All the characters came from Wingdings fonts, but one (A1) came from the Lucida Console font. Stimuli appeared on a white 12.7 × 10.16 cm note card. Each note card had a grid of nine rectangles, approximately 2.5 cm × 2.5 cm presented on one side. The sample stimulus appeared in the center grid, and the comparisons appeared in the outer four corners, randomized to vary the location of the correct comparison. Note cards were numbered on the back during the experiment to track presentation order.

Procedure. Participants were assigned randomly to one of two conditions: CARO present with 10 participants or CARO absent with 8 participants. Participants in both groups were trained and tested using the same protocol and symbols; the only difference between the two groups was the presence of a CARO displayed along with class-based stimuli on the note cards. Participants were instructed to match by circling the correct comparison with a pen that was supplied by the experimenter. Those in the CARO group read the following directions:

> The following is a test of matching characters. The format is known as matching to sample. In this test, a series of symbols appear on note cards.

The sample is presented in the center of the box. Answer choices will appear at the corners of each box. Your task is to match the sample in the center by circling the appropriate choice in one of the four corners. On every matching question, one possible answer will be "CA," which stands for "can't answer." This response can be used when the sample does not match any of the other responses. Here are some sample matching questions using letters.

The instructions then provided two examples, both using upper- and lowercase letters in a boxed format, demonstrating how to respond on trials and when it would be appropriate to respond with CA. Participants in the control (No-CARO) group received the following instructions lacking any mention of responding with CARO:

The following is a test of matching characters. The format is known as match to sample. In this test, a series of symbols appear on note cards. The sample symbol is presented in the center of the box. Answer choices will appear at the corners of each box. Your task is to match the sample in the center by circling the appropriate choice in one of the four corners. Here is a sample matching question using letters.

They then saw two boxed examples, one with upper- and lowercase letters and the other with characters (⊠, ¶, ©, and ۞), both displaying comparisons in three corners with one corner empty. The characters' example was accompanied by "For the second example, none of the responses clearly match ⊠." Additional instruction would be needed to respond, such as "When ⊠ is the sample, circle ¶. This instruction demonstrates how to correctly match the symbols."

The final paragraph in the two groups' directions read: "When you are ready to begin, the experimenter will present you with instructions on how to make matches in this experiment. Once the test starts, you will not be able to ask any questions regarding the format of the experiment, so please ask them now." Before each training phase, the experimenter presented participants with instructions that illustrated the baseline relations as described above. For both groups, the instructions read, for example, "When ϕ is the sample, circle ♑," illustrating each pair of stimuli depicted in Figure 1.

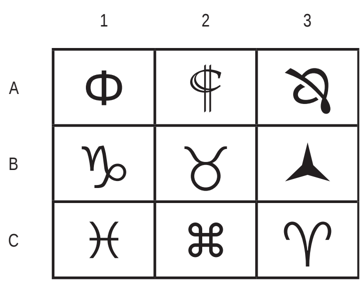

Figure 1. Stimuli used in Experiment 1. Symbols are characters from MS Word.

Table 1 presents the sequence of trained and tested relations used for all participants, representing a simple-to-complex (STC) protocol (Imam, 2006). Participants had to respond correctly on a minimum of four consecutive relations before advancing to the next training block. Thus, when participants were unable to demonstrate a previously trained relation, they again received instructions on all the failed relation(s) until they could respond correctly on the relation. For example, participants in the first block (training A1-B1), would need to respond correctly to four consecutive A1-B1 trials before moving on to the A2-B2 training block. Additionally, to ensure that performance on prior relations had not deteriorated, participants had to demonstrate A1-B1 during subsequent training blocks. For example, after A2-B2 was trained, A1-B1 trials reappeared with A2-B2 trials in another block, and so on. This process was completed for each trained relation and ensured that by the end of training, participants mastered and maintained all baseline relations. After training each AB and BC relation, their respective BA and CB symmetry relations were tested. Upon completing the CB symmetry test, participants received an additional training block of all baseline relations before a final test block that included the baseline trials and all derived symmetry, transitivity, and equivalence tests intermixed in a single test block. The final test block contained one trial of each baseline and symmetrical relation and six trials each of transitivity and equivalence relations.

Table 1
Sequence of Training and Testing Blocks Implemented in Experiment 1

Step	Block	Relation(s)	Consecutive Correct Responses Required to Advance
1	Training	A1-B1	4
2	Training	A2-B2	4
3	Training	Mixed A1-B1, A2-B2	5
4	Training	A3-B3	4
5	Training	Mixed A1-B1, A2-B2, A3-B3	6
6	Testing	B1-A1, B2-A2, B3-A3	—
7	Training	B1-C1	4
8	Training	B2-C2	4
9	Training	Mixed A1-B1, A2-B2, A3-B3	6
10	Training	Mixed B1-C1, B2-C2	5
11	Training	B3-C3	4
12	Training	Mixed B1-C1, B2-C2, B3-C3	6
13	Testing	C1-B1, C2-B2, C3-B3	—
14	Training	Mixed AB, BC	8
15	Testing	Mixed A1-B1, A2-B2, A3-B3; B1-C1, B2-C2, B3-C3; B1-A1, B2-A2, B3-A3; C1-B1, C2-B2, C3-B3; A1-C1, A2-C2, A3-C3; C1-A1, C2-A2, C3-A3	—

Results and Discussion

Figure 2 shows that mean percentage correct on the baseline relations during testing were as comparably high as those in training for the two groups, with 97% accuracy for the CARO group and 96% for the No-CARO group. The differences in performance between the two groups on baseline relations were not statistically significant, $t(16) = .493$, $p = .629$ for the original baseline; $t(16) = .239$, $p = .814$ for the tested baseline relations. Thus, both groups had fairly high accuracy rates on baseline relations during training and testing, an important factor in determining whether they had difficulty establishing the requisite baseline relations for the emergent equivalence relations that followed. Only one (10%) of the 10 participants in the CARO group, however, demonstrated equivalence, compared to six (75%) of the eight participants in the No-CARO group, despite equally high performances on baseline relations during training: 98% and 99%, respectively.

Performance on all derived relations was consistently lower for the CARO group compared to the No-CARO group. On the individual symmetry tests that followed training blocks, the CARO group had an accuracy of 30% whereas the No-CARO group had an accuracy of 88%; the difference was statistically significant, $t(16) = 3.00$, $p = .007$. Similarly, performances on the symmetry relations in the final mixed-test block indicated significantly lower accuracy for the CARO group (25%) compared to the No-CARO group (83%), as shown in Figure 2. Accuracy on transitivity trials was 10% for the CARO group compared to 85% for the No-CARO group; on equivalence trials, accuracy was 10.6% and

75%, respectively, for the two groups. A 2 (CARO availability: Yes or No) × 4 (relational type: B, S, T, and E) analysis of variance (ANOVA) conducted using GraphPad Prism version 5.04 for Windows (GraphPad, 2010) on the mixed-test data confirm these differences, showing statistically significant main effects of CARO availability, $F(1, 64) = 360950$, $p < .0001$, $\eta^2 = .47$; relational type, $F(3, 64) = 89785$, $p < .0001$, $\eta^2 = .35$; and their interaction, $F(3, 64) = 43728$, $p < .0001$, $\eta^2 = .17$; all Bonferroni pairwise posttest comparisons of differences in accuracy were statistically significant at $p < .0001$ (see Table 2 for obtained differences).

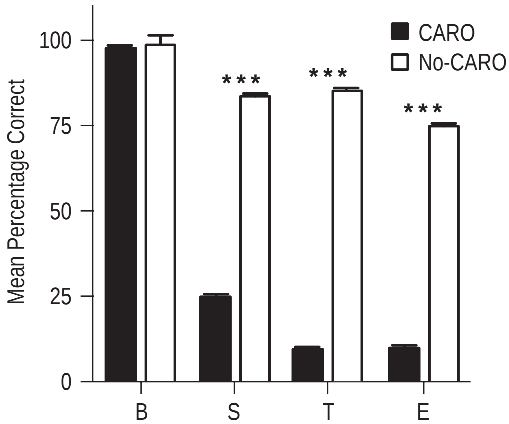

Figure 2. Mean and SD of percentage accuracy as a function of relational types from the final mixed test in Experiment 1 for the CARO and the No-CARO groups. Relational types included baseline (B), symmetry (S), transitive (T), and equivalence (E) relations.

*** $p < .0001$.

Table 2
Pairwise Comparisons of Differences in Accuracy on Relational Types During the Mixed-Test Blocks in Experiment 1

	Symmetry (S)		Transitivity (T)		Equivalence (E)	
	CARO	No-CARO	CARO	No-CARO	CARO	No-CARO
B	−70	−12.5	−86.7	−10.4	−86.1	−20.8
S			−16.7	2.1	−16.1	−8.3
T					0.6	−19.4

Note. B = Baseline.

On the whole, participants in the CARO group failed on most untrained relations, especially on transitivity and equivalence. Only two (20%) and three (30%) participants, respectively, passed the BA and CB symmetry tests in the CARO group compared to six (75%) and seven (88%), respectively, in the No-CARO group. Only one (10%) participant in the CARO group responded correctly on every transitive and equivalence trial, compared to six (75%) and five (63%) participants, respectively, in the No-CARO group. These results are fairly consistent with those of Duarte et al. (1998), who reported that one or two (3%–5%) CARO participants passed the symmetry, transitivity, and equivalence tests compared to 23 to 29 (47%–60%) of the No-CARO participants of the relevant, comparable (no instruction or baseline) groups (see Duarte et al., 1998).

The differences in performance between the two groups in Experiment 1 are attributable largely to CARO availability, as in the Duarte et al. (1998) study, being the only distinguishing feature of the two groups. When CARO was available, seven (70%) participants consistently selected CARO on all equivalence trials and eight (80%) selected CARO on all transitivity trials. Two participants primarily selected the CARO in addition to incorrect comparisons. Only one of the 10 participants in the CARO group correctly responded to all equivalence trials. Thus, of the group's 360 equivalence and transitive trials, only eight trials (0.02%) produced responses to incorrect comparisons. Notably, participants in the CARO group did not simply respond incorrectly to symmetry, transitivity, or equivalence trials during the mixed test; they overwhelmingly and consistently selected CARO. In fact, for trials on these emergent relations, only 2 (20%) of the 10 CARO-group participants selected any class-inconsistent responses, which constituted less than 1% of all derived relations test trials.

Despite the similarities between the present results and those of Duarte et al. (1998), some questions remain concerning the criterion used (four consecutive correct for each baseline relation) and the specific lack of training in the two studies, nonetheless. An important limitation was the lack of explicit training, which emanated from a reliance on participants' preexperimental history. While the CARO may inhibit class-based comparison selections to untrained relational trials, it is possible that this is due to the lack of explicit training needed to advance to the test phases. Another consideration is the type and frequency of feedback used. Feedback in this experiment, as in Duarte et al.'s study, came in the form of viewing illustrative instructions for a particular relation following incorrect trials for that relation. Consequently, prior to exposure to a particular relation's trial, participants would see the conditional instructions indicating the baseline relation. If the participant later responded incorrectly to that same relation, they were simply shown the directions again. Furthermore, participants who did not select incorrect responses on baseline trials received no feedback to that effect. An additional limitation is the present symbol set. The symbols were taken from fonts in Microsoft Office in order to provide participants with abstract stimuli. As the participants were college students, most of them familiar with this computer program, they may be familiar with the symbols prior to participation. Therefore, the symbols cannot be said to be completely unfamiliar for all participants.

Experiment 2

Having successfully replicated the broad effects of CARO on equivalence class formation reported by Duarte et al. (1998) using their original research design in Experiment 1, Experiment 2 examined the narrow question of whether those results with CARO availability could be replicated using a more standard equivalence training procedure (e.g., Innis et al., 1998). New sets of abstract and unfamiliar stimuli also were deployed.

Method

Participants. Fourteen students enrolled in an undergraduate, introductory summer debate course from John Carroll University participated in Experiment 2. They were at least 18 years of age and received extra credit for their participation.

Apparatus and materials. Participants worked on a Macintosh computer using the MTS software (Dube & Hiris, 1997) via a touch screen. Stimuli used in Experiment 2 appear in Figure 3. The sample always appeared in the center of the screen. When the participants responded on the sample, the comparisons appeared at the outer four corners of the screen, and their location was randomized so that the location of the correct responses varied. During training, when the CARO was not present, the fourth corner was blank. Correct comparison selections on trials with feedback produced a 1-s display of the word *correct* and a tone, and started a 1.5-s intertrial interval (ITI). Incorrect selections ended the trial, sounded a buzzer, darkened the screen for 1 s, and began the ITI. Participants were not informed of the letter and number designations of comparison stimuli and class membership

Procedure. Three sets of three-member classes were used. Participants experienced the different sets of stimuli using the same protocols for training and testing, except that "can't answer" was present as a fourth comparison during testing. There were no explicit instructions concerning the presence or absence of CARO or on how to use it, as was provided in Experiment 1. The following instructions were read to the participants at the start of the experiment:

> On the center of the screen, you will see a symbol. This will serve as the sample. When you press the symbol in the center, comparison symbols appear in the four corners. Your task is to select the appropriate symbol from the corner that corresponds with the symbol in the center.

Following these instructions, participants completed a pretraining block during which they matched uppercase letters (e.g., “A”) with their lowercase counterparts (e.g., “a”), to establish that they understood how the program worked. The experimenter demonstrated one trial performed correctly and one performed incorrectly to familiarize participants with the format and feedback conditions.

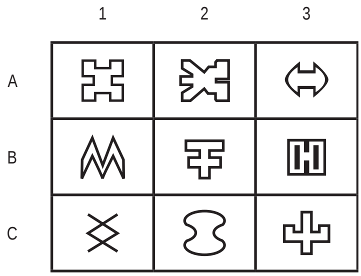

Figure 3. Stimuli used in Experiment 2 for the CARO conditions.

The sequence of presentation of blocks of trials followed that of Experiment 1 (see Table 1) based on Duarte et al. (1998). The criterion for advancing through training blocks was 12 correct consecutive trials. This ensured that participants had adequate exposure and demonstrated mastery of all relations within a training block, and thus CARO responding during testing could not be attributed to insufficient or inadequate training. Both the AB and BC training blocks utilized decreasing feedback of 100%, 50%, and 0%. Following a set of training blocks, symmetry blocks containing 6 trials for each relation for a total of 18 trials were implemented. Symmetrical, transitive, and equivalence relations were tested in separate blocks through which participants advanced regardless of their performance. In each of these blocks, each individual relation was tested six times. During the mixed-test block, all baseline and symmetrical relations appeared once, and the transitive and equivalence relations appeared six times.

Results and Discussion

Participants completed one to nine blocks of training of the AB, BC, and the combined AB-BC trials with the three levels of feedback. All participants met the requirement of 12 consecutive correct trials during training. As shown in Table 3, which presents the percentage accuracy data for each participant on the various test blocks, only 21% (3 of 14) of the participants demonstrated equivalence formation, scoring at least 90% (Lai, Made, and Isis); one other participant, Kwan, scored 50% on his individual transitivity block but did as well as the other three participants on the remaining test blocks. These results contrast sharply with those reported previously in studies using the STC protocol as in Experiment 2, showing better than 95% equivalence yields (e.g., Adams, Fields, & Verhave, 1993). That the yields in Experiment 2 are this low can be attributed to the adoption of CARO by most of the participants.

Figure 4 presents the total number of trials on which participants responded correctly (Corr), incorrectly (Incorr), or used CARO from the serially presented individual test blocks (I) and from the mixed test blocks (M). Five participants (Alex, Isis, Lai, Mariam, and Sore) did not use CARO at all on any relations. Whereas all but two, Phillip and Alex,

Table 3
Percentage Accuracy on Test Blocks for Individual Participants in Experiment 2

	Test Block				
	Symmetry		Transitivity	Equivalence	
Participant	BA	CB	AC	CA	Mixed Test
Lai	**100**	**100**	**100**	**100**	**100**
Made	**94**	**100**	**100**	**94**	**98**
Isis	**94**	**100**	**94**	89	**94**
Kwan	**100**	**100**	50	**94**	**92**
Mariam	94	100	100	67	83
Sore	100	94	72	72	42
Alex	94	94	33	33	38
Alyssa	100	100	6	6	23
Kyle	100	100	0	0	25
Lana	100	100	0	0	23
Tada	100	100	0	0	25
Tia	94	100	0	0	54
Rachel	94	89	0	0	25
Phillip	11	0	0	0	6

Note: Bold entries indicate equivalence formation.

of the 14 participants (86%) had at least 90% accuracy on symmetry trials from the individual test blocks, only three participants, Rachel, Sore, and Phillip, used CARO on these trials. Three participants (Isis, Lai, and Made) had at least 90% accuracy on both transitive and equivalence trials. In contrast, Kwan and Mariam met the same criterion on only equivalence and transitivity trials, respectively. In addition, whereas Kwan and Alyssa used CARO on 44% and 50%, respectively, on transitive trials, indicating 14% partial CARO selection, six participants (Kyle, Lana, Phillip, Rachel, Tada, and Tia) selected the CARO exclusively both on transitive and equivalence trials, representing 43% exclusive CARO usage. Taken together, then, 57% of the participants in the present experiment variously adopted the CARO, resulting in failures to demonstrate emergent equivalence relations prior to implementing the mixed test.

For the mixed-test blocks, Figure 4 shows that four participants (Isis, Kwan, Made, and Lai) had at least 90% accuracy on all relations including baseline, symmetry, transitive, and equivalence trials, signifying a 29% yield in successful equivalence formation despite the presence of CARO. This suggests that the availability of CARO does not necessarily mean it would control relational responding; these participants responded in a class-consistent manner despite CARO presence. Arguably, these participants simply ignored its presence, having learned specifically during training that there is always a correct choice on a given MTS task (Innis et al., 1998; Sidman, 1994), unlike in Duarte et al.'s (1998) study and Experiment 1, in which CARO was present during training as well as during testing. It remains unclear if additional kinds of history prompt a disregard of the "can't answer" response option when faced with abstract stimuli for the first time, in contrast to the familiar ones employed by Duarte et al. that yielded contrary outcomes.

Concerning the remaining (70%) participants' performance in the mixed tests, Figure 4 shows that eleven participants passed both AB and BC baseline trials; Sore passed all AB but missed two BC trials, whereas Alex and Phillip missed at least one of the baseline trials each on the AB and BC trials. On symmetry trials in the mixed test blocks, 12 participants passed on at least 2 of 3 BA or CB trials; only Sore and Phillip failed on these trials, the latter using CARO exclusively. Alex, Mariam, and Sore failed on transitivity and equivalence trials, representing 21% failure, without using CARO. The remaining eight participants used the CARO in various versions as depicted in Figure 4, indicating a 57% CARO adoption on transitivity and equivalence trials.

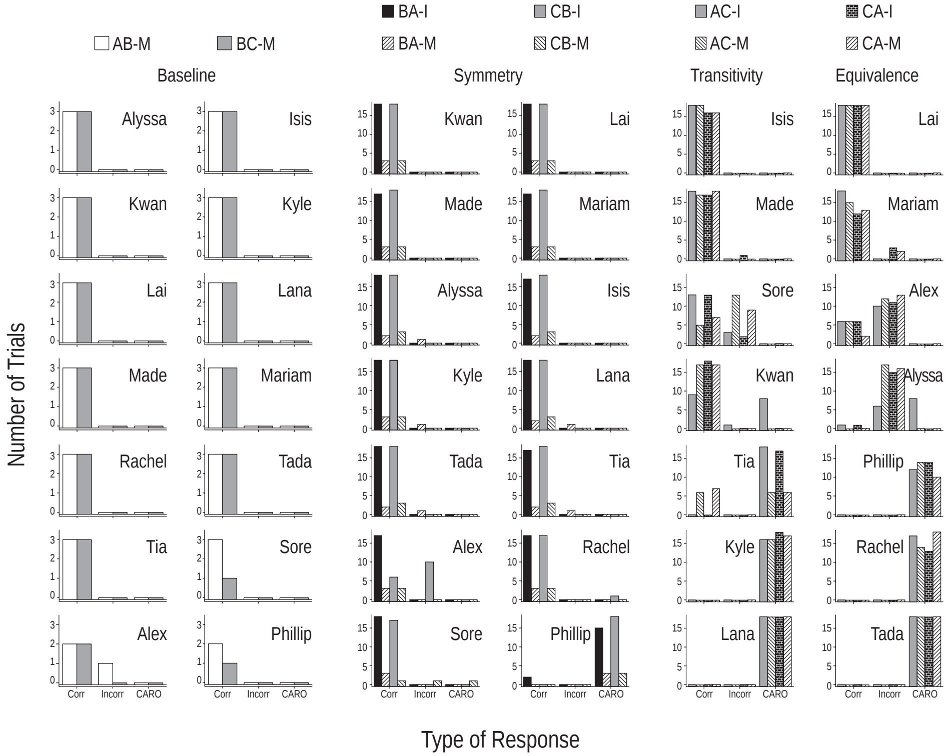

Figure 4. Number of trials on which participants chose correct (Corr), incorrect (Incorr), or the CARO comparisons on AB-BC baseline (left panels), BA-CB symmetry(middle panels), and AC transitivity and CA equivalence (right panels) trials in the individual (I) and the mixed (M) test blocks of Experiment 2.

Six of the 14 participants (43%) in Experiment 2 (Alyssa, Kyle, Lana, Rachel, Tada, and Tia) performed well on both baseline and symmetry trials in both the individual and mixed-test blocks but failed on both transitivity and equivalence trials. These patterns of relational-type responding, in which participants exhibit superior performances on baseline and symmetry relations and poor performances on transitive and equivalence relations, have been reported previously in the equivalence literature (e.g., Imam, 2001; Spencer & Chase, 1996), albeit not with respect to CARO usage. To the extent that participants used the CARO in Experiment 2, these results are consistent with those in Experiment 1 and those of other studies (e.g., Duarte et al., 1998; Innis et al., 1998); in Experiment 2, many participants relied on CARO, mostly on transitive and equivalence trials exclusively, and only a couple of them on symmetry trials, demonstrating amply the disruptive effects of the presence of the CARO on equivalence formation.

The noted 57% CARO adoption on transitivity and equivalence trials means that 14% of the participants failed to form equivalence relations without CARO usage. These failures may be a function of the LS training structure used by Duarte et al. (1998) and in the present study, in contrast to the sample-as-node (SAN) structure employed by Innis et al. (1998). Others have highlighted the difficulties typically encountered with this training structure (e.g., Imam, 2006; Saunders & Green, 1999). Although the presumed inferiority of LS training structures have been reported, there are a large number of demonstrations of the formation of three-member classes with LS training structures, where virtually all of the participants in a group form the classes with the exception of classes established using the simultaneous protocol (e.g., see Adams et al., 1993). As discussed further below, it could be that there is a combination of protocols and training structures that would interact with CARO availability to maximally impede equivalence formation.

The lack of exposure to the CARO prior to testing blocks in Experiment 2 parallels the no-training group from the Innis et al. (1998) study. Their no-training group did not involve the default response option until testing. Likewise, in Experiment 2, participants received no exposure to the CARO until during testing. In their initial equivalence testing, only 5 of the 16 participants (31.3%) demonstrated equivalence. Their results are comparable to our findings in which 4 of 14 participants (28.6%) demonstrated transitive and equivalence relations in the initial equivalence test blocks, and 4 participants (28.6%) demonstrated these relations in the final mixed-test block. These outcomes suggest that CARO availability during training may predispose some or more participants to their usage during testing, as they did in Duarte et al.'s study and in Experiment 1.

Some interesting data come from the participants with initially consistent responses who switched patterns during testing in Experiment 2. Two of these participants, Tia and Kwan, initially correctly identified the symmetry trials and then selected CARO on all transitive and equivalence trials. After the eighth transitive trial, as noted above, Kwan began selecting class-consistent comparisons on all trials except one, selecting 8 CARO, 1 incorrect response (A1-C3), and 9 class-consistent transitive responses. Kwan then correctly responded to all equivalence test trials and every trial in the final test block. Similarly, Tia selected CARO for all transitive and equivalence trials while correctly identifying symmetry relations in the individual test blocks. During the final test block, after selecting CARO on six transitive and seven equivalence trials, Tia began to select the correct responses for the remaining trials. Another participant, Phillip, correctly identified the first two BA symmetry trials before selecting CARO for every subsequent emergent-relations trial. These cases appear to represent delayed emergence of equivalence classes, which has been widely reported in the equivalence literature (e.g., Arntzen & Holth, 2000; Holth & Arntzen, 1998; Kato, de Rose, & Faleiros, 2008; Sidman, 1994).

General Discussion

To the extent that participants used the CARO in the two experiments of the present study, the results are consistent with the previous studies of nonresponse options (e.g., Duarte et al., 1998; Innis et al., 1998). Experiment 1 provided a direct systematic replication of Duarte et al.'s findings, based on a study of only two classes of familiar stimuli using a PAP format, by adopting similar techniques with abstract stimuli in three classes instead. As in the previous study, despite significant modifications in procedures, participants in the CARO group tended not to establish stimulus equivalence, compared to those in the No-CARO group. Experiment 2 offered an extension of these results to abstract, unfamiliar stimuli used in more standard stimulus-equivalence format, focusing on individual performance under more extensive training than was used by Duarte et al. and in Experiment 1. Many participants used the CARO mostly on transitive and equivalence trials than they did on symmetry trials. The preponderance of the evidence in the present study thus supports the growing literature on the disruptive effects of default-response options on equivalence formation (e.g., Duarte et al., 1998; Fields et al., 1993; Fields, Reeves, Adams, Brown, & Verhave, 1997; Innis et al., 1998; Reeve & Fields, 2001). In total, only five (21%) of the 24 participants who experienced CARO in the present study established equivalence classes, as 16 (67%) of them used CARO on the transitive and equivalence trials during the mixed test.

In Experiment 1, 90% of the CARO group used CARO on transitivity and equivalence trials compared to only 57% of participants in Experiment 2 of the present study. This relatively different impact of CARO availability on responses on these trials may be accounted for by several procedural differences in the two experiments. First, the experiments differed in the number of training trials used for the AB, BC baseline relations; whereas a minimum of four trials was imposed in Experiment 1, 12 consecutive correct responses were required in Experiment 2. In Experiment 1, there was a potential for differences in exposure to baseline relations for different participants regardless of CARO availability, as was possible in the Duarte et al. study as well. Inadequate and inconsistent experiences of requisite baseline relations for the emergence of equivalence therefore was a confound in the observed CARO usage because participants received no direct feedback on correct choices. This possibility was eliminated in Experiment 2 by increasing the total number of trials to criterion and providing direct feedback on all trials, hence minimizing substantial variability in the amount of training participants received on each relation. This difference between Experiment 2 and the Duarte et al. study and Experiment 1 of the present study may have contributed to diminished effect of CARO on transitivity and equivalence trials in the former compared to the latter experiments.

Second, the two experiments differed in terms of the type of stimuli used, unfamiliar but potentially accessible Microsoft Office characters in Experiment 1 and more abstract ones in Experiment 2. In this respect, however, the two experiments differed significantly from Duarte et al. (1998), who used familiar letters and numbers, thereby relying largely on participants' personal histories to establish baseline relations. By relying less on personal history and more on direct feedback, participants in Experiment 2 experienced greater control by baseline training, resulting in less reliance on the CARO. Third, they differed in their use of CARO during training and testing; in Experiment 1, like in Duarte et al.'s study, CARO was available during *both* training and testing, whereas in Experiment 2 it was available only during testing. As such, participants encountered the phrase "can't answer" for the first time during the serially presented individual test blocks in Experiment 2 of the present study (see also Innis et al., 1998). Based on the specific feedbacks in baseline training, participants might have learned to attend to class-based comparisons only in Experiment 2, having learned during training that there was always a correct answer, and therefore ignored the CARO more during testing compared to participants in Experiment 1 and in Duarte et al.'s study. Finally, whereas participants in Experiment 1 as in the Duarte et al. study received explicit instructions on the use of CARO, participants in Experiment 2 did not receive any instructions on availability or usage of the CARO.

The case of the four participants who experienced CARO but failed to use it in Experiment 2 appear to represent the kind of specific experimental control one would expect in the absence of extraexperimental history effects that are possible with human participants. Presumably, default-response options gain control when personal history overshadows experimental contingencies (see discussion below). One could speculate how that happens. The verbal nature of the typical default response would play a role, especially when abstract stimuli are deployed, in which case, say in a four-choice scenario where CARO or "none" option is available, the option may stand out and gain control over choices the participants make, unless instructions to the contrary are provided. Duarte et al.'s (1998) study provided a hint of this possibility when they used instructions in conjunction with the CARO. They found that more participants receiving nonrestrictive, in contrast to those receiving restrictive or no instructions, formed equivalence classes in the presence of the CARO. Future research on default responding might consider use of nonverbal default-response options for equivalence testing, perhaps akin to Innis et al.'s suggestion for "studies using an arbitrarily designated default-response option whose function is established entirely in the context of the experiment" (1998, p. 97).

When participants encounter the CARO, use it during tests for emergent relations, and thereby fail to demonstrate equivalence classes, how does one interpret such failures? For example, in Experiment 1, participants in the CARO group should be just as likely to form equivalence as their counterparts in the No-CARO group. The CARO availability may simply prevent some CARO participants from looking beyond any relation established during training; they simply select CARO instead of examining the comparisons for other possibilities and considering alternatives, much like the participants who received the restrictive instructions provided by Duarte et al. (1998). Indeed, Duarte et al. found that the restrictive-instruction group behaved much the same way as did the baseline group, with only 3%–5% of them demonstrating equivalence. Another possibility is that some participants may examine comparisons for other untrained relations, determine that these relations exist, but are hesitant to indicate them. Some evidence for this possibility in Experiment 1 is that, although five (50%) of the 10 CARO participants responded correctly on at least 1 trial of the emergent relations, only one of them (10%) consistently responded correctly on symmetry, transitive, and equivalence trials. The other four (40%) participants, at some point during testing, responded correctly on some trials of untrained relations but selected CARO for their other responses, demonstrating their recognition of these emergent relations, but choosing CARO anyway, upon testing for them. What, then, could account for such hesitation that results in the reluctance to select class-consistent comparisons and adopt the CARO instead?

If the CARO presence does not simply prevent equivalence class formation, as may have happened in Experiment 2, where no specific instructions about the availability or use of CARO were provided for participants, it could be that the classes did form but were overshadowed by the evocation of the CARO, preventing class-based choices. As noted in the introduction, Reeve and Fields (2001) reported such an effect with dimensionally defined stimulus classes in which variants failed to emerge as members of a perceptual class when “neither” was present as an option among class-based comparisons in their Experiment 1 but emerged successfully without it in their Experiment 2. The results suggest that the nonresponse option had overshadowed the emergence of class membership when it was present. A useful manipulation for future research would be to present CARO in tests of derived relations and then implement the test without CARO as in Reeve and Fields’ study, upon demonstrating failures on the tests with CARO.

If the interpretation of the Reeve and Fields’ (2001) study is valid, the question still remains as to what instigates the overshadowing effect of CARO. One possibility is that participants presented with nonresponse options simply ignore class-based comparison, perhaps because they need to get out of the experimental session as quickly as possible and an option such as CARO presents an easy way out; this, of course, would signify that weak experimental contingencies may be in place. Relying on personal history such as was the case in Experiment 1 and in the Duarte et al. (1998) study would therefore facilitate such behavior. Another possibility is that participants may recognize class-based choices but are hesitant to use them, perhaps, due to uncertainty. In that case, a participant may simply adopt “if not sure, use CARO” when CARO is presented as an option during tests of derived relations. Haste and uncertainty on the part of participants then might predispose them towards CARO adoption, thereby overshadowing class-consistent choices during testing. If haste makes CARO salient during testing, then perhaps a solution is to provide adequate or additional incentives to participants for their performances. After all, more often than not, it is common practice to use college-student volunteers in equivalence research, albeit for course credits. Additional or extensive training prior to tests of derived relations also might help reduce uncertainty during testing and therefore less reliance on CARO as evidenced by the results of Experiment 2 in the present study. The details and intricacies of how factors such as these might facilitate CARO usage remain subjects for further research.

When CARO is available and participants do not use it, or when CARO is not available at all but participants still fail to demonstrate equivalence classes, such failures cannot be attributed to the nonresponse option. In Experiment 1 of the present study, two participants (25%) did not respond in accord with equivalence, despite the nonavailability of CARO in the No-CARO group; in Experiment 2, three participants (21%) failed to demonstrate equivalence without adopting the CARO. These results suggest that the failures to establish stimulus equivalence in the present study as well as in the Duarte et al. (1998) study may not be wholly due to CARO availability. A common factor in the two studies is the LS training structure implemented, one that is reputed in the literature for low and inconsistent equivalence yields (e.g., Arntzen & Holth, 2000; Arntzen, Grondahl, & Eilifsen, 2010; Saunders, Chaney, & Marquis, 2005; Saunders & McEntee, 2004). A pertinent question then is whether the failures to establish equivalence in the present study, as in the Duarte et al. (1998) study, could be wholly due to the effect of CARO availability or partly due to the LS training structure deployed. It is possible that the difficulty engendered by the LS training structure contributes to the uncertainty a participant might experience leading to a consideration to use the CARO, having been ill-prepared in training for the requisite (AC) conditional discriminations needed for successful equivalence performance (see Saunders & Green, 1999). If this is the case, one way to address it would be to implement procedures such as those adopted previously by Saunders and McEntee (2004) to deal with similar issues (cf. Arntzen et al., 2010). Upon training the requisite AC conditional discriminations before tests for derived relations, Saunders and McEntee increased equivalence yields from 0% in Experiment 1 to 67% in Experiment 2! A similar strategy might reveal the relative contribution of LS to the disruptive effects of CARO availability during tests of emergent relations by using CARO with LS training structure in combination with AC training.

Besides training structure, another factor that has been implicated in differential equivalence yields in the literature is training protocol: the STC, the complex-to-simple (CTS), and the simultaneous protocols (SP; e.g., Adams et al., 1993; Arntzen et al., 2010; Imam, 2006), in that order of effectiveness in establishing equivalence. The prevailing protocol in the Duarte et al. (1998) and the present studies is the STC, immediately interspersing tests of derived relations with training of their relevant baseline relations (e.g., BA symmetry following AB training, etc.). As such, in terms of equivalence yields, these studies combined the best faring protocol with the worst faring training structure. The comparison of various studies of training structures provided by Saunders et al. (2005), which shows rather inconsistent findings on their effects on equivalence yield, suggest that considerations of the relative effects of CARO availability in the context of the various protocols and training structures is just beginning. Perhaps comparing CARO availability with nonavailability under, for example, the SP, say using three classes of four stimuli (to be consistent with these other studies), might reveal how disruptive of equivalence formation the CARO and other nonresponse options really are, or whether participants simply rely on them as face-saving options. Further research, therefore, is highly recommended.

As mentioned earlier in the introduction, various nonresponse options in a conditional discrimination context carry different implications. In comparing the results of Experiment 1 and Duarte et al.'s (1998) study, the CARO choice is equivalent to "neither" (see also Fields et al., 1993) in the latter case because participants had a binary context in which to make their choices. In Experiment 1, however, the CARO carried a different meaning, with participants' choices being in the context of three other comparison stimuli. Sidman (1980) and others (e.g., Imam, 2006) have discussed previously the problems associated with a two-choice conditional discrimination procedure. Applied to a default-response option, it is unclear yet how two- versus three-choice procedures may affect participants' performance. Is there a propensity for greater failure to form equivalence classes given a two versus three or more choices in conditional discriminations? The results of previous studies such, as Duarte et al. (1998) and Fields et al. (1993), both using two choices, compared to Innis et al. (1998) using three choices, suggest that there may be no differential outcome when a default response is present, notwithstanding the measurement issues associated with two-choice conditional discrimination (Sidman, 1980).

References

ADAMS, B. J., FIELDS, L., & VERHAVE, T. (1993). Effects of the test order on the establishment and expansion of equivalence classes. *The Psychological Record, 43*, 133–152.

ARNTZEN, E., GRONDAHL, T., & EILIFSEN, C. (2010). The effects of different training structures in the establishment of conditional discriminations and subsequent performance on tests for stimulus equivalence. *Psychological Record, 60,* 437–462.

ARNTZEN, E., & HOLTH, P. (2000). Probability of stimulus equivalence as a function of class size vs. number of classes. *Psychological Record, 50,* 79–104.

DUARTE, A. M., EIKESETH, S., ROSALES-RUIZ, J., & BAER, D. (1998). The effects of a can't-answer response option and instructions on stimulus equivalence. *Psychological Record, 48,* 631–646.

DUBE, W., & HIRIS, J. (1997). Match to Sample Program (Version 11.08a67) [Computer software]. Waltham, MA: E. K. Shriver Center for Mental Retardation.

EIKESETH, S., ROSALES-RUIZ, J., DUARTE, A., & BAER, D. (1997). The quick development of equivalence classes in a paper-and-pencil format through written instructions. *Psychological Record, 47,* 275–284.

FIELDS, L., ADAMS, B. J., BROWN, J. L., & VERHAVE, T. (1993). The generalization of emergent relations in equivalence classes: Stimulus substitutability. *Psychological Record, 43,* 235–254.

FIELDS, L., REEVE, K. F., ADAMS, B. J., BROWN, J. L., & VERHAVE, T. (1997). Predicting the extension of equivalence classes from primary generalization gradients: The merger of equivalence classes and perceptual classes. *Journal of the Experimental Analysis of Behavior, 68,* 67–91. doi: 10.1901/jeab.1997.68-67

GRAPHPAD. (2010). Prism for Windows (Version 5.04) [Computer software]. San Diego, CA: Author.

HOLTH, P., & ARNTZEN, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *Psychological Record, 48,* 81–110

IMAM, A. A. (2001). Speed contingencies, number of stimulus presentations, and the nodality effect in equivalence class formation. *Journal of the Experimental Analysis of Behavior, 76,* 264–288. doi: 10.1901/jeab.2001.76-265

IMAM, A. A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *Journal of the Experimental Analysis of Behavior, 85,* 107–124. doi: 10.1901/jeab.2006.58-04

INNIS, A., LANE S. D., MILLER, E. R., & CRITCHFIELD T. S. (1998). Stimulus equivalence: Effects of a default response option on emergence of untrained stimulus relations. *Journal of Experimental Analysis of Behavior, 70,* 87–102. doi: 10.1901/jeab.1998.70-87

KATO, O. M., DE ROSE, J. C., & FALEIROS, P. B. (2008). Topography of responses in conditional discrimination influences formation of equivalence classes. *Psychological Record, 58,* 245–267.

REEVE, K. F., & FIELDS, L. (2001). Perceptual classes established with forced-choice primary generalization test and transfer of function. *Journal of the Experimental Analysis of Behavior, 76,* 95–114. doi: 10.1901/jeab.2001.76-95

SAUNDERS, R. R., CHANEY, L., & MARQUIS, J. G. (2005). Equivalence class establishment with two-, three-, and four-choice matching to sample by senior citizens. *Psychological Record, 55,* 539–559.

SAUNDERS, R. R., & GREEN, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior, 72,* 117–137. doi: 10.1901/jeab.1999.72-117

SAUNDERS, R. R., & MCENTEE, J. E. (2004). Increasing the probability of stimulus equivalence with adults with mental retardation. *Psychological Record, 54,* 423–435.

SAUNDERS, R. R., & SHERMAN J. A. (1986). Analysis of the "discrimination-failure hypothesis" in generalized matching and mismatching behavior. *Analysis and Intervention in Developmental Disabilities, 6*, 89–107. doi: 10.1016/0270-4685(86)90008-x

SIDMAN, M. (1980). A note on the measurement of conditional discriminations. *Journal of the Experimental Analysis of Behavior, 33*, 285–289. doi: 10.1901/jeab.1980.33-285

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

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