Behavioural Processes 166 (2019) 103903

Contents lists available at ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

# Differential derived stimulus relations across probe-trial versus adduction testing are not a function of comparison-stimulus presentation

Adam H. Doughty*, Jenna A. Soydan

*Department of Psychology, College of Charleston, United States*

A R T I C L E I N F O

*Keywords:*
Adduction
Derived stimulus relations
Emergent learning
Stimulus equivalence
Human
Mouse click

A B S T R A C T

The present research assessed whether unreliable derivation of three-node equivalence relations in adduction testing is due to the absence of comparison-stimulus presentation. College students first learned arbitrary-matching-to-sample discriminations (AB, BC, CD, and DE) with Arabic numbers (A), nonsense syllables (B, C, D), and nonrepresentational forms (E) as stimuli. Participants then were exposed to either probe-trial testing or adduction testing. In probe-trial testing, each trial involved one of the four E stimuli (e.g., E1) as a sample and each of the four A stimuli (A1, A2, A3, A4) as comparisons. Successful performance involved participants selecting the corresponding A stimulus in the absence of differential consequences (e.g., A1 in the presence of E1). In adduction testing, each trial involved an arithmetic equation such that successful performance involved combining derived EA relations with simple math skills. Critically, unlike in typical adduction testing, participants were presented with four response options to select from (i.e., comparison stimuli). In probe-trial testing, each participant derived the EA relations, whereas no participant performed successfully in adduction testing. The present findings suggest that the unreliable derivation of three-node equivalence relations in adduction testing is not due to the typical absence of comparison stimuli. The present research challenges experimental psychologists to clarify the role of testing context in identifying the necessary and sufficient conditions that control emergent learning.

## 1. Introduction

Deriving stimulus relations is a powerful form of emergent learning (e.g., Barnes-Holmes et al., 2017; Critchfield et al., 2018); however, its necessary and sufficient conditions remain unclear (e.g., Barnes-Holmes et al., 2017; Doughty et al., 2018). Several investigations suggest that the testing for derived relations sometimes is necessary for the development of such learning (e.g., Doughty and Best, 2017; Doughty et al., 2018; Doughty et al., 2014a; Doughty et al., 2014b; Haimson et al., 2009; Sidman et al., 1985). Despite the aforementioned findings, as well as a growing literature involving different derived-relations assessments (e.g., Arntzen et al., 2017; Barnes-Holmes et al., 2006; Cummins et al., 2018; Dickins, 2015; Fields et al., 2014), few comparative analyses of assessment type on derived relations exist.

Adduction is the emergence of a novel and complex composite skill after its simpler, component skills have been learned (e.g., Andronis et al., 1997; Chase, 2003; Epstein, 1987). Three studies have investigated adduction with derived relations (Arntzen et al., 2015; Bucklin et al., 2000; Rippy and Doughty, 2017). Bucklin et al. first taught participants AB and BC relations where the A, B, and C stimuli were (previously learned) Hebrew symbols, nonsense syllables, and (previously learned) Arabic numbers, respectively. In adduction testing, arithmetic questions were posed using the Hebrew symbols. For example, A1 and A2 had to be added such that adduction required deriving the previously untested transitive AC relations and combining them with their extant math skills. Arntzen et al. extended these findings by demonstrating adduction involving relations other than transitive (i.e., symmetrical and equivalence); however, their participants derived these relations in probe-trial testing before the adduction assessment. Rippy and Doughty examined adduction involving previously untested equivalence relations. Group CA learned AB and BC relations (Arabic numbers [A], nonsense syllables [B], and nonrepresentational stimuli [C]), whereas Group EA learned AB, BC, CD, and DE relations (Arabic numbers [A], nonsense syllables [B, C, and D], and nonrepresentational stimuli [E]). In adduction testing, the C and E stimuli for Groups CA and EA, respectively, were presented such that participants had to combine simple math skills with previously untested equivalence relations with one (Group CA) or three (Group EA) nodes. Successful adduction occurred for each CA participant but with only one EA participant.

* Corresponding author at: Department of Psychology, College of Charleston, 57 Coming Street, Charleston, SC, 29424, United States.
*E-mail address:* doughtya@cofc.edu (A.H. Doughty).

https://doi.org/10.1016/j.beproc.2019.103903
Received 21 May 2019; Received in revised form 3 July 2019; Accepted 4 July 2019
Available online 05 July 2019

Identifying the variables responsible for the unreliable EA adduction in Rippy and Doughty (2017) should improve our understanding of how the testing context facilitates derived relations. The present experiment assessed derived EA relations across probe testing and modified adduction testing. In probe testing, some number of comparison stimuli surrounds a sample stimulus, and participants select a comparison. In adduction testing, participants are presented with a configuration that can be conceptualized as a sample without comparisons to select from. The present experiment explored the role of such choices by measuring derived EA relations across probe testing and adduction testing with comparisons.

## 2. Method

### 2.1. Participants

Eight College of Charleston students (six female and two male) between the ages of 18 and 22 participated. Participants were told the research would involve one 3-h laboratory visit, earning them approximately $30.00.

### 2.2. Apparatus

An 8 x 13.5 ft. room had four workstations separated by dividers. Each workstation had a desk and chair. Each desk had an iMac or eMac, keyboard (not used by the participants), and mouse. The contingencies were programmed and responses recorded using MTS version 11.6.7 (Dube, 1991). Adduction testing involved 12 x 7 cm. flashcards with nonrepresentational stimuli, numbers, and mathematical operations.

### 2.3. Procedure

Phase 1 had five conditions, and all participants were treated identically. They first read:

> "Welcome to our study! In this part of the study, you will work alone on the computer for several sessions. In each session, the computer will present you with many trials. On each trial, you will be presented with one item, click on that item and additional items will appear. Click the mouse over any one of the surrounding items that you think "goes with" the one in the center, and one of two events will occur: (1) a star will appear on the screen or (2) the screen will darken. If a star appears, then you were correct and earned money. If the screen darkens, then you were incorrect and did not earn money. Your task is to earn as much money as possible. Good luck!"

There were 96 trials in each session of Condition 1 wherein participants learned AB relations (see Table 1; the stimuli were identical to the EA group in Rippy and Doughty, 2017). The four AB (i.e., number to nonsense syllable) relations were presented such that each A sample occurred on 24 trials in each session. Across these 24 trials (e.g., A1), the correct comparison (e.g., B1) occurred in each screen corner six times. Every trial began with only a sample in the middle of the screen. After a click over it (observing response), the comparisons immediately appeared with the sample. The stimuli were pseudorandomly organized such that each sample could not occur on more than three consecutive trials, and the correct comparison could not occur in the same location on more than three consecutive trials. Clicking the correct comparison immediately resulted in stars on the screen for 1 s. Clicking an incorrect comparison resulted in a 1.5-s dark screen. A resetting intertrial interval (ITI) of 1.5 s was used wherein the screen was blank (i.e., white). Condition 1 continued for at least two sessions and until there were no more than two errors per discrimination in the last session. Conditions 2, 3, and 4 were identical to Condition 1 except that participants learned the BC, CD, and DE relations, respectively.

Each session in Condition 5 consisted of 192 trials. There were 48 trials each of the AB, BC, CD, and DE relations (12 trials with each sample). Across these 12 trials, the correct comparison occurred three times in each corner. Other procedural details (e.g., consequence delivery) remained unchanged. The condition continued for at least one session and until there were no more than two errors per discrimination in a session.

**Table 1**
Stimuli during training for both groups and examples of stimuli during adduction testing (note that each trial in adduction testing also presented four response options not illustrated here).

| **Training Stimuli** | | | | | **Adduction Stimuli** | |
|---|---|---|---|---|---|---|
| **A1** | **B1** | **C1** | **D1** | **E1** | |  |
| 2 | yok | fic | het |  | X | 4 |
| **A2** | **B2** | **C2** | **D2** | **E2** | | |
| 5 | mof | tep | sul |  | |  |
| **A3** | **B3** | **C3** | **D3** | **E3** | + | 6 |
| 7 | bal | loz | gix |  | | |
| **A4** | **B4** | **C4** | **D4** | **E4** | |  |
| 8 | kad | dut | roj |  | x | 5 |

In Phase 2, participants were randomly assigned to the probe-trial group or adduction group. In the Probe-trial group, each participant read:

> "In your next session, there never will be any stars or dark screen. The computer still will record whether your choice is correct or incorrect. You still will receive money for each correct choice. Good Luck!"

The session was constructed in a manner similar to Phase 1. One E stimulus always was the sample, and the four A stimuli always were the comparisons (i.e., 96 EA trials). Each E sample occurred on 24 trials with its corresponding comparison in each corner on six trials. Other than the elimination of consequences, other procedural details were unchanged (i.e., unsystematic placement of the incorrect comparisons, 1.5-s ITI, and pseudo-random trial presentation).

Each participant in the adduction group read:

> "Your next session will consist of me presenting you with 96 flashcards. Please calculate the answer and circle it. After you finish one card, I will hand you the next one. No feedback will be given during the session. However, after the session your answers will be assessed and money provided for each correct answer. Good luck!"

The flashcards were nearly identical to Rippy and Doughty (2017). The experimenter provided a pen and presented one flashcard at a time. Each completed flashcard was set to the side such that participants could not respond to previous flashcards. The time between flashcards was only as long as it took to place the previous flashcard to the side and present the next one (i.e., approximately a 1.5-s ITI). Each flashcard required participants to multiply (48 cards) or add (48 cards) an E stimulus (from Phase 1) with an Arabic number ranging from 1 to 12 (e.g., E1 + 9). Each E stimulus appeared on 24 flashcards (12 multiplication and 12 addition), and each Arabic number appeared on eight flashcards (four multiplication and four addition). The flashcards were organized pseudorandomly such that there were no more than three consecutive addition or multiplication problems, and no more than three consecutive flashcards with the same E stimulus. Importantly, and unlike Rippy and Doughty, each flashcard included four answers to select from (i.e., one "correct" comparison and three "incorrect" comparisons). The correct comparison was the number that resulted from a successful EA derivation (and accurate multiplication or addition),

**Table 2**
Accuracy scores (i.e., percent correct) for each participant in each session.

| Condition | Participants AD | JA | TD | VF |
|---|---|---|---|---|
| Adduction Group | | | | |
| AB | 80 | 94 | 98 | 89 |
| | 100 | 99 | 100 | 98 |
| BC | 90 | 88 | 93 | 84 |
| | 100 | 100 | 100 | 100 |
| CD | 76 | 95 | 73 | 90 |
| | 99 | 100 | 99 | 100 |
| DE | 97 | 95 | 97 | 84 |
| | 100 | 100 | 100 | 100 |
| AB – DE | 100 | 100 | 100 | 100 |
| Adduction | 27 | 23 | 24 | 16 |
| | CB | CC | IG | MG |
| Probe-trials Group | | | | |
| AB | 82 | 77 | 99 | 62 |
| | 100 | 100 | 100 | 99 |
| BC | 85 | 86 | 96 | 89 |
| | 98 | 100 | 100 | 100 |
| CD | 85 | 81 | 84 | 88 |
| | 97 | 99 | 99 | 100 |
| DE | 97 | 94 | 99 | 96 |
| | 99 | 99 | 100 | 100 |
| AB – DE | 99 | 100 | 100 | 100 |
| Probe Trials | 62 | 93 | 85 | 94 |

whereas the incorrect comparisons were the numbers that would result from inconsistent EA derivations (e.g., deriving A2, A3, or A4 from E1).

## 3. Results

Table 2 displays session-by-session accuracy scores. Each participant completed Phase 1 in the minimum number of sessions. Performance differed across groups in Phase 2. Each participant in the Adduction group had an accuracy score lower than 30% (i.e., at chance levels), whereas three of the participants in the Probe-trials group had accuracy scores at or above 85% (Participant CB scored 62%). Mean accuracies were 22.40% and 83.50% in the Adduction and Probe-trials groups, respectively, which were significantly different: $t(6) = 7.801$, $p = 0.0002$.

## 4. Discussion

The derivation of EA relations was markedly different across the two assessments, despite all participants learning the baseline discriminations rapidly and comparably. This differential derivation occurred even with adduction trials presented in a sample-and-comparison format. Trial presentation was electronic in probe testing and manual in adduction testing. This factor seems insignificant because previous adduction studies with manual presentation have obtained successful performance (Arntzen et al., 2015; Bucklin et al., 2000; Rippy and Doughty, 2017). Instead, the critical factor seems to be the type of derived relations involved. Successful adduction has included previously untested, one-node transitive relations (Bucklin et al., 2000) as well as previously tested (Arntzen et al., 2015), and previously untested (Rippy and Doughty, 2017), one-node equivalence relations. Unsuccessful adduction has involved previously untested, three-node equivalence relations (Rippy and Doughty, 2017; the present research). Thus, participants are less likely to adduce derived relations as they become more "difficult" (e.g., greater nodal distance; e.g., Barnes-Holmes et al., 2017; Doughty et al., 2018). Importantly, the present research suggests that this differential derivation is attributable to some factor(s) unrelated to the presence versus absence of comparison stimuli per se.

The present research has implications for identifying the experiences accrued in typical training and testing that underlie derived relational learning. The stimulus configuration in probe-trial testing resembles discrimination training. One stimulus appears in the location designated as sample surrounded by multiple stimuli designated as comparisons. Prior to testing, participants have an immediate history of selecting among current comparisons as a function of the current sample, independent of other factors (e.g., previous sample, comparison configuration). This history involves learning that only one comparison is correct in the presence of only one sample and that one of the current comparisons always is correct. The stimulus configuration in adduction differs considerably from training. Additional stimuli are present (e.g., an Arabic number and term related to the necessary operation [e.g., a "+"]), and the participant typically is not presented with stimuli to select from (except in the present experiment). Instructions typically describe a "simple" response selection (probe trials) or (for adduction) a "calculation" (requiring multiple steps). If successful derivation is conceptualized as a composite skill, there are fewer component skills that must be occasioned in probe testing. The antecedent stimuli needed to occasion the component skills are more contiguous (temporally and spatially) in probe testing. The stimuli also occur in locations that recently have been correlated with particular stimulus functions (as samples and comparisons) and after the aforementioned learning history (that one and only one current comparison is related to the current sample).

The role of assessment type must be understood as investigators refine their approaches to derived relations (e.g., Barnes-Holmes et al., 2017; Doughty et al., 2018). Performance sometimes differs as a function of assessment type, but sometimes only under particular circumstances (e.g., with more complex relations).

## References

Andronis, P.T., Layng, T.J., Goldiamond, I., 1997. Contingency adduction of "symbolic aggression" by pigeons. Anal. Verbal Behav. 14, 5–17. https://doi.org/10.1007/BF03392913.

Arntzen, E., Petursson, P.I., Sadeghi, P., Eilifsen, C., 2015. Equivalence class formation in accuracy or speed conditions: immediate emergence, adduction, and retention. The Psych. Record. 65, 141–159. https://doi.org/10.1007/s40732-014-0097-9.

Arntzen, E., Granmo, S., Fields, L., 2017. The relation between sorting tests and matching-to-sample tests in the formation of equivalence classes. The Psych. Record. 67, 81–96. https://doi.org/10.1007/s40732-016-0209-9.

Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., Stewart, I., 2006. Do you really know what you believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a direct measure of implicit beliefs. Ir. J. Psychol. Med. 32, 169–177.

Barnes-Holmes, D., Barnes-Homes, Y., Luciano, C., McEnteggart, C., 2017. From the IRAP and REC model to a multi-dimensional multi-level framework for analyzing the dynamics of arbitrarily applicable relational responding. J. Contex. Behav. Sc. 6, 434–445. https://doi.org/10.1016/j.jcbs.2017.08.00.

Bucklin, B.R., Dickinson, A.M., Brethower, D.M., 2000. A comparison of the effects of fluency training and accuracy training on application and retention. Perform. Improv. Quart. 13, 140–163. https://doi.org/10.1111/j.1937.8327.2000.tb00180.x.

Chase, P.N., 2003. Behavioral education: pragmatic answers to questions of novelty and efficiency. In: Lattal, K.A., Chase, P.N. (Eds.), Behavior Theory and Philosophy. Plenum/Kluwer, New York, NY, pp. 347–367.

Critchfield, T.S., Barnes-Holmes, B., Dougher, M.J., 2018. What Sidman did: historical and contemporary significance of research on derived stimulus relations. Perspect. Behav. Sc. 41, 9–32. https://doi.org/10.1007/s40614-018-0154-9.

Cummins, J., Roche, B., Tyndall, I., Cartwright, A., 2018. The relationship between differential stimulus relatedness and implicit measure effects sizes. J. Exp. Anal. Behav. 110, 24–38. https://doi.org/10.1002/jeab.437.

Dickins, D.W., 2015. Vocalizing phonologically correct non-word stimuli during equivalence class formation. Europ. J. Behav. Anal. 16, 248–278. https://doi.org/10.1080/15021149.2015.1083284.

Doughty, A.H., Best, L., 2017. Transfer of function and prior derived-relations testing. Behav. Proc. 143, 4–6. https://doi.org/10.1016/j.beproc.2017.07.010.

Doughty, A.H., Brierley, K.P., Eways, K.R., Kastner, R.M., 2014a. Effects of stimulus discriminability on discrimination acquisition and stimulus-equivalence formation: assessing the utility of a multiple schedule. The Psych. Rec. 64, 287–300. https://doi.org/10.1007/s40732-014-0001-7.

Doughty, A.H., Leake, L.W., Stoudemire, M.L., 2014b. Failure to observe untested derived stimulus relations in extinction: implications for understanding stimulus-equivalence formation. J. Exp. Anal. Behav. 102, 311–326. https://doi.org/10.1002/jeab.111.

Doughty, A.H., Brenner, S.E., Fox, M.L., Rippy, S.M., 2018. Emergence of simpler untested derived stimulus relations in extinction: implications for understanding derived relational learning. The Psych. Rec. 68, 49–60. https://doi.org/10.1007/s40732- 018-

0267-2.
Dube, W.V., 1991. Computer software for stimulus control research with Macintosh computers. Exp. Anal. Hum. Behav. Bull. 9, 28–30.
Epstein, R., 1987. The spontaneous interconnection of four repertoires of behavior in a pigeon (Columba livia). J. Comp. Psychol. 101, 197–201. https://doi.org/10.1037/0735-7036.101.2.197.
Fields, L., Arntzen, E., Moksness, M., 2014. Stimulus sorting: a quick and sensitive index ofequivalence class formation. The Psych. Rec. 64, 487–498. https://doi.org/10.1007/s40732-014-0034-y.
Haimson, B., Wilkinson, K.M., Rosenquist, C., Ouimet, C., McIlvane, W.J., 2009. Electrophysiological correlates of stimulus equivalence processes. J. Exp. Anal. Behav. 92, 245–256. https://doi.org/10.1901/jeab.2009.92-245.
Rippy, S.M., Doughty, A.H., 2017. Adduction of untested derived stimulus relations depends on environmental complexity. Behav. Proc. 143, 1–3. https://doi.org/10.1016/j.beproc.2017.07.008.
Sidman, M., Kirk, B., Wilson-Morris, M., 1985. Six-member stimulus classes generated by conditional-discrimination procedures. J. Exp. Anal. Behav. 43, 21–42. https://doi.org/10.1901/jeab.1985.43-21.