CrossMark

Learning to write without writing: Writing accurate descriptions of interactions after learning graph-printed description relations

Jack Spear[2] · Lanny Fields[1,2]

Abstract Interpreting and describing complex information shown in graphs are essential skills to be mastered by students in many disciplines; both are skills that are difficult to learn. Thus, interventions that produce these outcomes are of great value. Previous research showed that conditional discrimination training that established stimulus control by some elements of graphs and their printed descriptions produced some improvement in the accuracy of students' written descriptions of graphs. In the present experiment, students wrote nearly perfect descriptions of the information conveyed in interaction-based graphs after the establishment of conditional relations between graphs and their printed descriptions. This outcome was achieved with the use of special conditional discrimination training procedures that required participants to attend to many of the key elements of the graphs and the phrases in the printed descriptions that corresponded to the elements in the graphs. Thus, students learned to write full descriptions of the information represented by complex graphs by an automated training procedure that did not involve the direct training of writing.

The ability to provide complete and accurate descriptions of complex situations or information is a skill that is essential to function effectively in today's complex world. For example, an individual should be able to explain the circumstances that precede an illness, an economic calamity, or a car accident. In addition, students of any natural science or social science such as psychology, sociology, or economics should be able to write complete and accurate descriptions of graphs that depict the interactive effects of two or more academically relevant variables on a phenomenon of interest in these domains. For example, in the realm of psychology, a therapist should be able to provide complete and accurate descriptions of the circumstances that precede the occurrence the problem behaviors exhibited by his/her clients. In another example, a student majoring in psychology should be able to write descriptions of the interactive effects of two variables on behavior (Behavior Analyst Certification Board 2012; Fields et al. 2009; Garfield and Ahlgren 1988; Garfield and Chance 2000; Mulhern and Wylie 2004). These skills, however, are generally regarded as being most difficult to learn. Thus, the development of efficient and reliable interventions that address this issue would be of substantial academic value (Garfield and Ahlgren 1988). Such interventions could also be used to establish analogous repertoires in many of the domains mentioned above.

The matter can be addressed in the context of learning to write paragraph-length descriptions of the interactive effects of two independent variables on some behavior. Figure 1 shows representations of each of four different types of interaction, with each on a different row. Each row contains a graph that represents one type of interaction and a printed paragraph that contains a full and accurate description of the information presented in the graph. From top to bottom, the graphs and printed texts in a row represent no interaction (NO), a crossover interaction (XOVR), a divergent interaction (DVR), and a synergistic interaction (SYN).

✉ Jack Spear
jspear@gradcenter.cuny.edu

1   Queens College of the City University of New York, Flushing, NY, USA

2   The Graduate Center of the City University of New York, New York, NY, USA

#1

NO Interaction [Co+ +]

#5. Depression score was a direct function of antidepressant dosage for late risers. In addition, depression score was a direct function of antidepressant dosage for early risers. For each level antidepressant dosage, the depression score was an inverse function of rising time. Increasing the antidepressant dosage produced a constant difference in depression scores for late risers relative to early risers. Depression score functions for late- and early-risers did not intersect at any level of antidepressant dosage.

#2

CROSSOVER Interaction [Co+ +]

#6. For high school graduates, the number of hours slept was an inverse function of the number of hours exercised per day. For college graduates, the number of hours slept was a direct function of the number of hours exercised per day. For different numbers of hours spent exercising, then, education reversed the effect of time spent exercising on the number of hours slept. The hours slept functions for individuals of different education levels intersected at an intermediate level of hours spent exercising.

#3

DIVERGENT Interaction [Co+ +]

#7. The anxiety level was a direct function of days of yoga per week for low-income individuals, and an inverse function of days of yoga per week for high-income individuals. For each number of days of yoga per week, the anxiety level was greater for low- than high-income individuals. Income level reversed the directional effect of days of yoga on anxiety level. Finally, the anxiety level functions for high- and low-income individuals did not intersect at any level of days of yoga.

#4

SYNERGISTIC Interaction [Co+ +]

#8. Pulse rate was a direct function of salt consumption for graduate students. In addition, pulse rate was a direct function of salt consumption for college students. For each level of salt consumption, type of school attended produced higher pulse rates for graduate students relative to college students. Increasing the amount of salt eaten produced a growing difference in pulse rate for graduate students relative to college students. Pulse rate functions for graduate and college students did not intersect at any level of salt consumption.
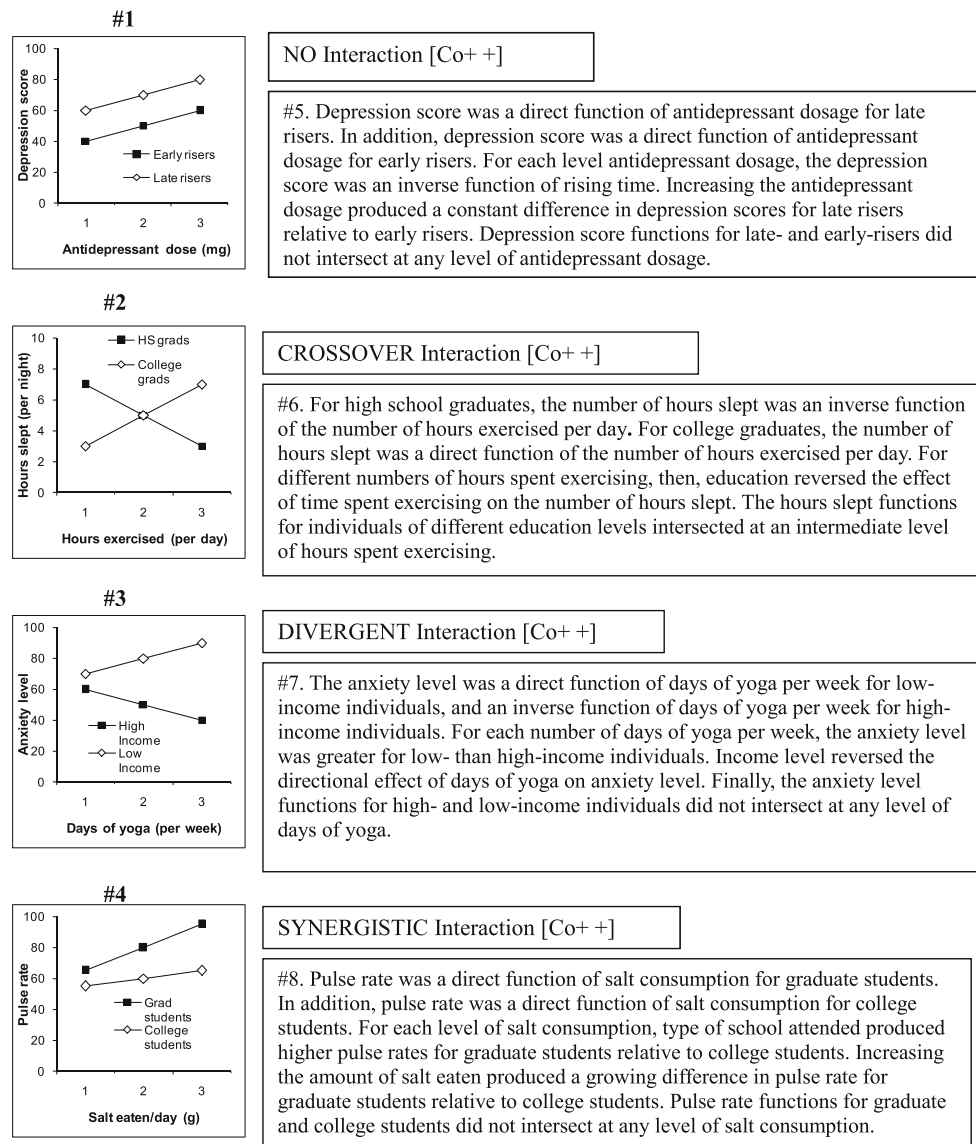
Fig. 1 Graphs depicting interactions and their corresponding accurate printed descriptions

One way to induce the writing of complete and accurate description of the information presented in each graph would be to establish conditional relations between a graph and its corresponding printed paragraph (Graph → Printed Description). Assuming a pre-existing generalized text-copying repertoire, which can be considered to be a complex response (Skinner 1957), and that the participants are attending to all of the elements of the material to be copied, the presentation of the printed text in the graph-text relations should occasion the copying of the printed text. In addition, it has been very extensively documented that a response made in the presence of one member of a conditional relation or an equivalence class typically transfers to the other members of that relation or class (Belanich and Fields 2003; Cowley et al. 1992; Eikeseth and Smith 1992; Fields et al. 1996; Groskreutz et al. 2010; Perez-Gonzalez et al. 2008; Rosales et al. 2011;

Sidman 1971; Sidman and Cresson 1973; Walker and Rehfeldt 2012; Walker et al. 2010). It follows logically that these two processes, along with the establishment of graph-printed text relations, should result in the writing of paragraph-length descriptions of the interactive effects of two variables on behavior when an individual is shown graphs that depict interactions. In addition, this should occur in the absence of direct training of the writing behavior.

Spear and Fields (submitted) evaluated this analysis and found that a very small proportion of college students who received such conditional discrimination training wrote accurate and complete descriptions; instead, the bulk of participants wrote descriptions that were very incomplete, inaccurate, or both. Indeed, using other sets of complex stimuli, other researchers found similar negative results (Walker et al. 2010; Walker and Rehfeldt 2012). In one recent experiment, however,

accurate multiword descriptions emerged when extensive prompting was added to the writing test (Sella et al. 2014).

Walker and Rehfeldt (2012) noted that the stimuli used as the members of the conditional relations were complex, but the selection of the correct comparison stimuli did not require attending to all of the elements of the stimuli during conditional discrimination training, as is the case with most "traditional" conditional discrimination training. The production of accurate responding in the transfer tests requires a participant to include references to all or many of the features, or elements, of the complex stimuli used during conditional discrimination training. Because traditional conditional discrimination training does not necessarily establish control by all of these elements of the sample and comparison stimuli, a participant would not have learned to attend to all of the elements needed to construct a complete and accurate written response.

This interpretation of failed performances prompted Spear and Fields (submitted) to devise a conditional discrimination training procedure that differed from the traditional formats used by Walker and Rehfeldt (2012) and that required participants to attend to the many elements of the complex sample and comparison stimuli and also the correspondence between the elements of those sample and comparison stimuli. When the participants learned Graph-printed description conditional relations in this manner, a moderate proportion of participants wrote reasonably accurate descriptions of graphs that depicted the interactive effects of two variables on behavior. These results, then, supported the general validity of the theoretical analysis presented by Walker and Rehfeldt (2012). In addition, the fact that Spear and Fields (submitted) did not find maximal yields (i.e., the production of accurate descriptions by all participants) implies that some of the key elements of the stimuli used in training were not being attended to and thus could not influence the writing of accurate and complete descriptions of interaction-indicative graphs.

The present experiment had two goals: first, to identify all of elements of graphs that depict interactions and the corresponding elements of their printed descriptions, and second, to identity some combinations of these graphic and textual elements that would produce highly accurate written descriptions of graphicly represented interactions by most students.

Elements in graphs and printed descriptions of interactions

Each interaction-based graph depicted in Fig. 1 contains five non-relational elements and two relational elements (Fields and Spear 2012). The non-relational elements are the two independent variables, one on the abscissa that is designated IV1, and the other in the legend that is designated IV2; the dependent variable that is designated DV; the directionality of

the function on a figure that has the higher y-intercept and is designated Dir Hi; and the directionality of the function on the graph with the lower y-intercept and is designated Dir Lo. All of these elements are considered to be non-relational because each describes a single element in a graph.

The two relational elements in the graphs are referred to as such because they denote the interactive effects of the two independent variables on behavior. Both are defined by reference to the two functions that are present in each graph. The first relational element is referred to as Interact, because it reflects how changes in the magnitude of one independent variable modulate the effect of the values of the other independent variable on behavior (Fields and Spear 2012). The other relational element is the possible intersection of the two functions because it defines the point of contact of one function relative to the other. This relational element is designated as Intersect. These seven elements characterize all interaction-indicative graphs, and thus the four in Fig. 1.

Table 1 is a glossary of the terms mentioned above and also lists their abbreviations, definitions, and usages. Each of these elements in a graph also has a descriptive textual counterpart. These are represented in the paragraphs that are to the right of each graph in Fig. 1 and are also listed in Table 1. Each non-relational element is described by one phrase or term. There are two sets of textual terms that are counterparts of the relational term Interact that describe the interactions. One set includes the words *produced* and *reversed*, and is referred to as Interact:p , and the other set of terms includes the words *constant* or *growing difference*, and is referred to as Interact:d. Finally, the textual element that is the counterpart of the relational term Intersect is indicated by the phrases *does/does not intersect*. All of these terms are also listed in Table 1.

Sources of control by graphic and textual elements of interactions

Theoretically, the writing of complete and accurate written descriptions of each type of interaction should require the prior establishment of control by graphic and textual representations of all of the relational and non-relational elements listed in Table 1. This means that the relational elements in all graph-text relations have to control behavior in the context of all of the non-relational elements for all four types of interaction. One way of codifying all of these nested stimulus control relations is illustrated in Table 2.

Traditional conditional discrimination

To begin, the graphs that represent the four types of interaction have to be discriminated from each other. This can be done by traditional conditional discrimination training, where the graphs are presented as sample stimuli with printed

Table 1 A glossary of terms and abbreviations used to denote the elements of graphs and their accompanying printed descriptions. Examples refer to the features of a graph depicting no interaction, as shown in the top row of Fig. 1. The relevant terms of each sentence are italicized

Abbreviation	Name	Definition	Example
IV1	Independent variable 1	Independent variable listed on the abscissa	Antidepressant dose
IV2	Independent variable 2	Independent variable listed in the legend	Rising time
Dir Lo	Directionality of the lower function	Slope of function with the lower y-intercept	Depression score was a *direct* function….
Dir Hi	Directionality of the higher function	Slope of function with the higher y-intercept	Depression score was a *direct* function….
Interact:p	Interaction: produced or reversed	*Produced* or *reversed* in the interaction statement	Rising time *produced* a constant difference…
Interact:d	Interaction: difference	*Difference* term in the interaction statement	Rising time produced a *constant* difference….
Intersect	Intersection of the functions	*Intersection* term of the intersection statement	Depression scores *did not intersect*….
DV	Dependent variable	Variable listed on the ordinate	Depression score

descriptions of each type of interaction presented as a set of comparisons. As mentioned above, this ensures discrimination of the graphs but does not ensure control by the all of the elements of the graphs or the elements of the printed descriptions. Finally, these conditional relations must be established for all four types of interaction. This level of stimulus control is called Traditional conditional discrimination.

Targeted conditional discriminations

The second level of control involves the establishment of conditional discriminations between one relational element and its corresponding textual equivalent, of which there are three. The first is between the interaction element of a graph and one set of textual terms, or Interact:d. The second uses the

Table 2 All possible conditional stimulus control relations among the relational and non-relational elements, and those established during the intervention components. An "x" indicates the training of the form of stimulus control listed in the leftmost column, while a "–" indicates an untrained stimulus control relation

Intervention components

Train typeq	Element	x-x-x	W&R	S&F	A	B	A + B
Traditional CD		-	x	x	x		x
Targeted CD							
	Interact:d	-	-	x	x	x	x
	Interact:p	-	-	-	-	x	x
	Intersect	-	-	-	-	x	x
Decontextualized control of Interact:d with non-relational elements							
	Interact:d & IV1	-	-	x	x		x
	Interact:d & IV2	-	-	x	x		x
	Interact:d & Dir Lo	-	-	x	x		x
	Interact:d & Dir Hi	-	-	x	x		x
Decontextualized control of Interact:p with non-relational elements							
	Interact:p & IV1	-	-	-	-	-	-
	Interact:p & IV2	-	-	-	-	-	-
	Interact:p & Dir Lo	-	-	-	-	-	-
	Interact:p & Dir Hi	-	-	-	-	-	-
Decontextualized control of Intersect with non-relational elements							
	Intersect & IV1	-	-	-	-	-	-
	Intersect & IV2	-	-	-	-	-	-
	Intersect & Dir Lo	-	-	-	-	-	-
	Intersect & Dir Hi	-	-	-	-	-	-
	DV	-	-	-	-	-	-

Note: The + s and –s in each column indicate the conditional discriminations that were and were not trained, respectively. The W&R column indicates procedures used by Walker and Rehfeldt (2012). The S&F column indicates procedures used by Spear and Fields (submitted). See text for more information

same graphic element but in combination with another set of textual terms, or Interact:p. The third is between the point of intersection or a lack thereof and its corresponding textual terms, or Intersect. All of these targeted conditional relations are established for all four types of interaction and in the context of only one set of non-relational terms. For example, Interact:p would be established in the context of the same set of specific IV1, IV2, Dir Hi, Dir Lo, and DV values. This would also be done for each of the four types of interaction.

Decontextualized targeted conditional discriminations

Each targeted conditional discrimination is established in the context of only one set of non-relational elements. The third level of control involves decontextualizing each of the targeted conditional discriminations by the re-establishment of each in the presence of all of the non-relational elements. Thus, each targeted conditional discrimination would have to be established with each of the four non-relational elements: IV1, IV2, Dir Hi, and Dir Lo. Each is represented in one of three lower the sections of Table 2.

Summary

According the analysis presented above and symbolized by Table 2, there are 16 stimulus-stimulus or stimulus-stimulus-stimulus relations that must be established to ensure that a participant is attending to all of the stimulus elements in graph-printed text relations for one type of interaction. Since there are four types of relations, attending to all graphic and textual elements involves the establishment of 64 conditional relations.

Predictive analysis of prior research

Table 2 places into a quantitative perspective the outcomes of the experiments reported by Walker and Rehfeldt (2012) and Spear and Fields (submitted). Walker and Rehfeldt (see W&R column in Table 2) used Traditional conditional discrimination training procedures alone, which did not ensure the establishment of any targeted conditional relations or decontextualized targeted conditional discriminations. Thus, the absence of most of the precursor relations would predict very poor writing performances. Spear and Fields (submitted; see S&F column in Table 2) established the traditional discriminations along with one of the three targeted conditional discriminations listed in Table 2, and one set of decontextualized conditional discriminations (Interact:d) in the context of IV1, IV2, Dir Hi, and Dir Lo. Since there are many untrained relations that are needed to ensure attention to all of the elements in the graphs and printed descriptions, these procedures were expected to produce a level of writing

accuracy that was modest but still higher than that obtained by Walker and Rehfeldt. The data supported this prediction.

Emergence of accurate descriptions of interactions

The present experiment determined how the writing of accurate descriptions of interaction-based graphs could be maximized by the establishment of some of the forms of stimulus control that are listed in Table 2. Four groups were studied, each of which had three intervention conditions. After each, a test was administered in which the participant was asked to write descriptions of four graphs that depicted different types of interaction. In the x-x-x condition, no interventions separated each test, and a minimal level of writing accuracy was expected. In the x-A-B condition, participants received an A component that contained traditional discrimination training, one form of targeted conditional discrimination training that established Interact:d relations with one set of non-relational elements, and the re-establishment of the Interact:d relations in the context of IV1, IV2, DIR:Hi, and DIR:Lo. This was then followed by a B component that contained two other forms of targeted conditional discrimination training that established Interact:p relations and Intersect relations. Because this combination of A-training and B-training required participants to attend to all of the relational and non-relational elements in the graphs and printed descriptions, we predicted a very high level of accuracy in the written descriptions of the data represented by graphs showing the interactive effects of two variables on behavior.

Finally, two additional groups were studied to isolate the effects of the stimulus control topographies established in A- and B-training in isolation. Participants in the x-A-A and x-x-B conditions received only A- or B-training, respectively. Because each of these conditions involved the training of only some of the conditional discriminations in Table 2, we expected each to produce writing accuracy that was substantially less than that produced by the training of A and B together.

Method

Participants

All participants were undergraduates students enrolled in a course in Introductory Psychology, and were drawn from the Psychology 101 subject pool at Queens College, City University of New York. Participants received course credit for the first training session, and were paid US$50 if they completed a follow-up training session. A total of 52 participants were recruited, and were assigned to each of four groups in a block-randomized basis without replacement, resulting in 13 per group. Participants who learned the conditional

discriminations trained in session 1 were invited to participate in session 2, which was held within seven days of session 1. A participant was not invited to participate in session 2 if she/he did not acquire all of the trained relations in session 1. Each session lasted from 1.5 to 2 hours maximum, with actual session duration determined by participant performance.

Materials

Setting, hardware and software Each participant worked independently in a cubicle measuring 2.5 m × 2.5 m. During training, all stimuli were presented on a PC-based desktop computer. Stimuli were presented in an RGB computer monitor with a 19-in diagonal dimension and a 4 × 3 width to height ratio. Custom software developed in our laboratory was used to conduct this trial-based experiment. The software controlled the presentation of all stimuli in the experiment and collected information about all of stimuli presented in each trial and all responses made to the stimuli.

Stimuli The experiment involved the administration of a writing-test booklet, and of the stimuli listed in Fig. 1 and Tables 3, 4, 5, 6, 7, 8, and 9, which were used in Traditional conditional discrimination training, Targeted conditional discrimination training, and Decontextualized non-relational training.

The writing-test booklets used for each of the three writing tests were identical and consisted of four 8.5 × 11-in pages that were stapled together in the upper left corner. The top of each page contained a graph of one of the four types of interaction, as shown on the left side of Fig. 1. The following instruction was printed above each graph: "Please write a complete and accurate description of the following graph." The space

beneath the instruction was provided for the writing of the description and constituted about two-thirds of the page.

Traditional discrimination training involved the presentation of the graphs and printed descriptions depicted in Fig. 1. These stimuli were also used as the samples and positive comparisons in all of the remaining conditional discriminations established in the experiment.

Tables 3, 4, 5, 6, 7, 8, and 9 contain the negative comparisons used in other components of the experiment. The negative comparisons used to establish the targeted conditional discriminations Interact:d, Interact:p, and Intersection are presented in Tables 3, 4, and 5, respectively.

The negative comparisons used to established decontextualized targeted conditional discriminations using IV1, IV2, Dir Lo, and Dir Hi for no, crossover, divergent, and synergistic interactions are presented in Tables 6, 7, 8, and 9, respectively. The stimuli in Tables 3, 4, 5, 6, 7, 8, and 9 and in Fig. 1 are identified by unique numbers. Those numbers are used to indicate their inclusion in each stage of training as outlined in Tables 10 and 11. In addition, the meanings of the labels attached to each stimulus will be described below.

Procedure

Experimental design The experiment was conducted using four groups, each of which experienced three stages and is named after these stages: x-x-x, x-A-A, x-x-B, and x-A-B. The first stage in each group involved no intervention (denoted by the first "x" in the string used to denote a group) followed by a writing pre-test. These test performances provided a baseline measure of the accuracy of writing descriptions of graphs that depicted the interactive effects of two variables on behavior in the absence of any experimental intervention.

Table 3 The stimuli used as incorrect comparisons to establish targeted conditional control by Interact:d (Id-) for each of the four types of interaction. The incorrect words in each description are bolded. This bolding did not appear during training

No	XOVR	DVR	SYN
#9. Depression score was a direct function of antidepressant dosage for late risers. In addition, depression score was a direct function of antidepressant dosage for early risers. For each level of antidepressant dosage, the depression score was an inverse function of rising time. Increasing the antidepressant dosage produced a **growing** difference in depression scores for late risers relative to early risers. Depression score functions for late- and early-risers did not intersect at any level of antidepressant dosage.	#10. For high school graduates, the number of hours slept was an inverse function of the number of hours exercised per day**.** For college graduates, the number of hours slept was a direct function of the number of hours exercised per day. For different numbers of hours spent exercising, then, education produced a **constant** difference in the effect of time spent exercising on the number of hours slept. The hours slept functions for individuals of different education levels intersected at an intermediate level of hours spent exercising.	#11. The anxiety level was a direct function of days of yoga per week for low-income individuals, and an inverse function of days of yoga per week for high-income individuals. For each number of days of yoga per week, the anxiety level was greater for low- than high-income individuals. Income level produced a **constant** difference in the directional effect of days of yoga on anxiety level. Finally, the anxiety level functions for high- and low-income individuals did not intersect at any level of days of yoga.	#12. Pulse rate was a direct function of salt consumption for graduate students. In addition, pulse rate was a direct function of salt consumption for college students. For each level of salt consumption, type of school attended produced higher pulse rates for graduate students relative to college students. Increasing the amount of salt eaten produced a **constant** difference in pulse rates for graduate students relative to college students. Pulse rate functions for graduate and college students did not intersect at any level of salt consumption.

Table 4 The stimuli used as incorrect comparisons to establish targeted conditional control by Interact:p (Ip-) for each of the four types of interaction. The incorrect words in each description are bolded. This bolding did not appear during training

No	XOVR	DVR	SYN
#13. Depression score was a direct function of antidepressant dosage for late risers. In addition, depression score was a direct function of antidepressant dosage for early risers. For each level of antidepressant dosage, the depression score was an inverse function of rising time. Increasing the antidepressant dosage **did not produce** a constant difference in depression scores for late risers relative to early risers. Depression score functions for late- and early-risers did not intersect at any level of antidepressant dosage.	#14. For high school graduates, the number of hours slept was an inverse function of the number of hours exercised per day. For college graduates, the number of hours slept was a direct function of the number of hours exercised per day. For different numbers of hours spent exercising, then, education **did not reverse** the effect of time spent exercising on the number of hours slept. The hours slept functions for individuals of different education levels intersected at an intermediate level of hours spent exercising.	#15. The anxiety level was a direct function of days of yoga per week for low-income individuals, and an inverse function of days of yoga per week for high-income individuals. For each number of days of yoga per week, the anxiety level was greater for low- than high-income individuals. Income level **did not reverse** the directional effect of days of yoga on anxiety level. Finally, the anxiety level functions for high- and low-income individuals did not intersect at any level of days of yoga.	#16. Pulse rate was a direct function of salt consumption for graduate students. In addition, pulse rate was a direct function of salt consumption for college students. For each level of salt consumption, type of school attended produced higher pulse rates for graduate students relative to college students. Increasing the amount of salt eaten **did not produce** a growing difference in pulse rates for graduate students relative to college students. Pulse rate functions for graduate and college students did not intersect at any level of salt consumption.

The second stage varied across groups and involved no intervention for two groups (x-x-x and x-x-B) or A training for two groups (x-A-B and x-A-A). In all groups, these conditions were followed by a second administration of the writing test. Differences in the writing tests after the completion of stage 2 would indicate how the emergence of accurate written descriptions of the data portrayed in the interaction graphs was influenced by the effects of test repetition alone (in the x-x-x and x-x-B groups) and by A training alone (in the x-A-A and x-A-B groups).

The third stage involved another no-intervention condition in the in the x-x-x group, re-administration of A training (in the x-A-A group), or administration of B training (in the x-x-B and x-A-B groups). In all cases, these conditions were followed by a third and last administration of the writing test. Differences in the writing tests at the end of stage 3 would indicate how the emergence of accurate written descriptions of the data portrayed in the interaction graphs was influenced by test repetition alone (in the x-x-x group), by the overtraining of the A conditional relations (in the x-A-A group), by B training alone (in the x-x-B group), and by the combination of A and B training (in the x-A-B group). A comparison of the final performances of the three training groups, then, would indicate the relative effects of A and B training alone, as well as their

Table 5 The stimuli used as incorrect comparisons to establish targeted conditional control by Intersect (Is-) for each of the four types of interaction. The incorrect words in each description are bolded. This bolding did not appear during training

No	XOVR	DVR	SYN
#17. Depression score was a direct function of antidepressant dosage for late risers. In addition, depression score was a direct function of antidepressant dosage for early risers. For each level of antidepressant dosage, the depression score was an inverse function of rising time. Increasing the antidepressant dosage produced a constant difference in depression scores for late risers relative to early risers. Depression score functions for late- and early-risers **intersected** at an intermediate level of antidepressant dosage.	#18. For high school graduates, the number of hours slept was an inverse function of the number of hours exercised per day. For college graduates, the number of hours slept was a direct function of the number of hours exercised per day. For different numbers of hours spent exercising, then, education reversed the effect of time spent exercising on the number of hours slept. The hours slept functions for individuals of different education levels **did not intersect** at any level of hours spent exercising.	#19. The anxiety level was a direct function of days of yoga per week for low-income individuals, and an inverse function of days of yoga per week for high-income individuals. For each number of days of yoga per week, the anxiety level was greater for low- than high-income individuals. Income level reversed the directional effect of days of yoga on anxiety level. Finally, the anxiety level functions for high- and low-income individuals **intersected** at an intermediate level of days of yoga.	#20. Pulse rate was a direct function of salt consumption for graduate students. In addition, pulse rate was a direct function of salt consumption for college students. For each level of salt consumption, type of school attended produced higher pulse rates for graduate students relative to college students. Increasing the amount of salt eaten produced a growing difference in pulse rate for graduate students relative to college students. Pulse rate functions for graduate and college students **intersected** at an intermediate level of salt consumption.

Table 6 Printed descriptions used to establish targeted conditional control by the non-relational elements of graphs and descriptions (Co + −) depicting no interaction (Interact: NO). The incorrect words in each description are bolded. This bolding did not appear during training

Depression score was a direct function of **hours exercised** for late risers. In addition, depression score was a direct function of **hours exercised** for early risers. For each level of **hours exercised**, the depression score was an inverse function of rising time. Increasing the number of **hours exercised** produced a constant difference in depression scores for late risers relative to early risers. Depression score functions for late- and early-risers did not intersect at any level of **hours exercised**.	#21 + Id, Ip, Is -- IV1
Depression score was a direct function of antidepressant dosage for **high school graduates**. In addition, depression score was a direct function of antidepressant dosage for **college graduates**. For each level of antidepressant dosage, the depression score was an inverse function of **education**. Increasing the antidepressant dosage produced a constant difference in depression scores for **high school-** relative to **college graduates.** Depression score functions for **high school** and **college graduates** did not intersect at any level of antidepressant dosage.	#22 + Id, Ip, Is -- IV2
Depression score was a direct function of antidepressant dosage for late risers. In addition, depression score was an **inverse** function of antidepressant dosage for early risers. For each level of antidepressant dosage, the depression score was an inverse function of rising time. Increasing the antidepressant dosage produced a constant difference in depression scores for late risers relative to early risers. Depression score functions for late- and early-risers did not intersect at any level of antidepressant dosage.	#23 + Id, Ip, Is -- Dir Lo
Depression score was an **inverse** function of antidepressant dosage for late risers. In addition, depression score was a direct function of antidepressant dosage for early risers. For each level of antidepressant dosage, the depression score was an inverse function of rising time. Increasing the antidepressant dosage produced a constant difference in depression scores for late risers relative to early risers. Depression score functions for late- and early-risers did not intersect at any level of antidepressant dosage.	#24 + Id, Ip, Is -- Dir Hi

combined effects (A + B). In all groups, the first and second writing tests were conducted in session 1 and the third writing test was conducted in session 2.

Trial structure, contingencies, and feedback All training was conducted using trials that contained one graph as a sample stimulus along with two, three, or four written descriptions of graphs as the comparison stimuli. The sample stimulus (a graph) was located in the top middle of the screen with the comparison stimuli below the sample stimulus. If there were two comparison stimuli, they were located below and to the left and right of the sample; if there were three, they were located below and to the left, right, and directly below the sample; if there were four, they were located equally spaced below the sample. The comparison stimuli were presented

with their positions randomized across trials. In addition, the letters 'A,' 'B,' 'C,' and 'D' were presented, one each, below the comparisons and their positions were also randomized with respect to each comparison. Thus, there was no correlation between a comparison paragraph and the 'A,' 'B,' 'C,' or 'D' letters used as comparison labels. Prior to training, the following instructions were presented to participants:

In this experiment, you will be presented with graphs and descriptions.
A graph will appear on the top of the screen, and descriptions will appear on the bottom.
Press the key on the keyboard that corresponds to the description that best matches the graph: either A, B, C, or D.

Table 7 Printed descriptions used to establish targeted conditional control by the non-relational elements of graphs and descriptions (Co + −) depicting crossover interaction (Interact: XOVR). The incorrect words in each description are bolded. This bolding did not appear during training

For high school graduates, the number of hours slept was an inverse function of **antidepressant dosage.** For college graduates, the number of hours slept was a direct function of **antidepressant dosage**. For different levels of **antidepressant dosage**, then, education reversed the effect of **antidepressant dosage** on the number of hours slept. The hours slept functions for individuals of different education levels intersected at an intermediate level of **antidepressant dosage**.	#25 + Id, Ip, Is -- IV1
For **late risers**, the number of hours slept was an inverse function of the number of hours exercised per day**.** For **early risers**, the number of hours slept was a direct function of the number of hours exercised per day. For different numbers of hours spent exercising, then, **rising time** reversed the effect of time spent exercising on the number of hours slept. The hours slept functions for individuals of different **rising times** intersected at an intermediate level of hours spent exercising.	#26 + Id, Ip, Is -- IV2
For high school graduates, the number of hours slept was an inverse function of the number of hours exercised per day**.** For college graduates, the number of hours slept was an **inverse** function of the number of hours exercised per day. For different numbers of hours spent exercising, then, education reversed in the effect of time spent exercising on the number of hours slept. The hours slept functions for individuals of different education levels intersected at an intermediate level of hours spent exercising.	#27 + Id, Ip, Is -- Dir Lo
For high school graduates, the number of hours slept was a **direct** function of the number of hours exercised per day**.** For college graduates, the number of hours slept was a direct function of the number of hours exercised per day. For different numbers of hours spent exercising, then, education reversed the effect of time spent exercising on the number of hours slept. The hours slept functions for individuals of different education levels intersected at an intermediate level of hours spent exercising.	#28 + Id, Ip, Is -- Dir Hi

Call the experimenter if you have any questions.
If not, press "Enter" to continue.

During trials that were followed by informative feedback, selection of the correct comparison was immediately followed by presentation of the word "Right!" and selection of an incorrect comparison was immediately followed by presentation of the word "Wrong!" Participants were then required to press either the 'R' or 'W' key, respectively, to progress to the next trial. When informative feedback was not provided, selection of any comparison was immediately followed by presentation of the words "No Feedback," and participants were required to press the 'E' key to progress to the next trial.

All trials were of unlimited duration. Each required a response to a comparison stimulus and then to the feedback stimulus. No participants had to be dismissed from the experiment for not meeting these response requirements.

Establishment of conditional discriminations

The two training procedures, A and B training, involved a mix of Traditional conditional discrimination training, one of three forms of Targeted conditional discrimination training, and decontextualized and targeted conditional discrimination training. A and B training are described below in separate sections.

A-training The A-training protocol involved the establishment of the Traditional, targeted, and decontextualized non-relational forms of conditional stimulus control listed in the A column of Table 2. The sample and comparison stimuli used to establish these sources of control are elaborated in Table 10.

Rows 1–4 of Table 10 represent trials used to establish traditional, graph-text conditional discriminations. This procedure will be illustrated with reference to row number 1.

Table 8 Printed descriptions used to establish targeted conditional control by the non-relational elements of graphs and descriptions (Co + −) depicting divergent interaction (Interact: DVR). The incorrect words in each description are bolded. This bolding did not appear during training

The anxiety level was a direct function of **salt consumption** for low-income individuals, and an inverse function of **salt consumption** for high-income individuals. For each level of **salt consumption**, the anxiety level was greater for low- than high-income individuals. Income level reversed the directional effect of **salt consumption** on anxiety level. Finally, the anxiety level functions for high- and low-income individuals did not intersect at any level of **salt consumption**.	#29 + Id, Ip, Is -- IV1
The anxiety level was a direct function of days of yoga per week for **graduate students**, and an inverse function of days of yoga per week for **college students**. For each number of days of yoga per week, the anxiety level was greater for **graduate students** than for **college students**. **Type of school attended** reversed the directional effect of days of yoga on anxiety level. Finally, the anxiety level functions for **graduate and college students** did not intersect at any level of days of yoga.	#30 + Id, Ip, Is -- IV2
The anxiety level was a direct function of days of yoga per week for low-income individuals, and a **direct** function of days of yoga per week for high-income individuals. For each number of days of yoga per week, the anxiety level was greater for low- than high-income individuals. Income level reversed the directional effect of days of yoga on anxiety level. Finally, the anxiety level functions for high- and low-income individuals did not intersect at any level of days of yoga.	#31 + Id, Ip, Is -- Dir Lo
The anxiety level was an **inverse** function of days of yoga per week for low-income individuals, and an inverse function of days of yoga per week for high-income individuals. For each number of days of yoga per week, the anxiety level was greater for low- than high-income individuals. Income level reversed the directional effect of days of yoga on anxiety level. Finally, the anxiety level functions for high- and low-income individuals did not intersect at any level of days of yoga.	#32 + Id, Ip, Is -- Dir Hi

During these trials, a graph sample stimulus, such as graph number 1 in Fig. 1, was presented with two printed description comparisons. One was a complete and accurate description of the graph (i.e., the Co++, comparison number 5 in Fig. 1), and another was a completely inaccurate description of the sample graph (the Co− −, comparison number 6 in Fig. 1), which was also a complete and accurate description of one of the three other graphs. Because the Co++ and Co− − differed with respect to all elements, correct responding required attention to and acquisition of control by only one element of the graph and only one element of a printed description. Thus, this procedure did not insure attention to all of the elements in the graphs and their corresponding printed descriptions. Similarly

structured trials were used to establish traditional graph-printed description relations for the other three types of interaction. These trials are elaborated in rows 2–4 of Table 10.

After the establishment of these traditional conditional discriminations, targeted conditional control by Interact:d was established for each of the four types of interaction. Rows 5–8 of Table 10 represent the trials used in this training. These trials each featured the presentation of a graph sample stimulus, along with two comparisons: the Co++ and Co − +. The Co+ + was an accurate description of the graph sample, and contained accurate references to all relational and non-relational elements of the graph. On the other hand, the Co − + contained an inaccurate reference to one of the relational

Table 9 Printed descriptions used to establish targeted conditional control by the non-relational elements of graphs and descriptions (Co + −) depicting synergistic interaction (Interact: SYN). The incorrect words in each description are bolded. This bolding did not appear during training

Pulse rate was a direct function of **days of yoga** per week for graduate students. In addition, pulse rate was a direct function of **days of yoga** per week for college students. For each number of **days of yoga**, type of school attended produced higher pulse rates for graduate students relative to college students. Increasing the number of **days of yoga** produced a growing difference in pulse rate for graduate students relative to college students. Pulse rate functions for graduate and college students did not intersect at any number of **days of yoga.**

#33
+ Id, Ip, Is
-- IV1

Pulse rate was a direct function of salt consumption for **low-income individuals**. In addition, pulse rate was a direct function of salt consumption for **high-income individuals**. For each level of salt consumption, **income level** produced higher pulse rates for **low-income** relative to **high-income individuals**. Increasing the amount of salt eaten produced a growing difference in pulse rate for **low-** relative to **high-income individuals**. Pulse rate functions for **low-** and **high-income individuals** did not intersect at any level of salt consumption.

#34
+ Id, Ip, Is
-- IV2

Pulse rate was a direct function of salt consumption for graduate students. In addition, pulse rate was an **inverse** function of salt consumption for college students. For each level of salt consumption, type of school attended produced higher pulse rates for graduate students relative to college students. Increasing the amount of salt eaten produced a growing difference in pulse rate for graduate students relative to college students. Pulse rate functions for graduate and college students did not intersect at any level of salt consumption.

#35
+ Id, Ip, Is
-- Dir Lo

Pulse rate was an **inverse** function of salt consumption for graduate students. In addition, pulse rate was a direct function of salt consumption for college students. For each level of salt consumption, type of school attended produced higher pulse rates for graduate students relative to college students. Increasing the amount of salt eaten produced a growing difference in pulse rate for graduate students relative to college students. Pulse rate functions for graduate and college students did not intersect at any level of salt consumption.

#36
+ Id, Ip, Is
-- Dir Hi

elements–Interact:d—of the graph sample, and accurate references to all of the non-relational elements. This comparison is abbreviated Id-. Consistent correct responding would thus require control by Interact:d of graphs and descriptions of each of the four types of interaction, as a participant could distinguish Co++ from Co − + by attending to the relational, but not the non-relational, elements of graphs and descriptions. These trials thus established control by Interact:d.

Once mastered, training to establish control by the four non-relational elements: IV1, IV2, Dir Lo, and Dir Hi, began. These trials are represented by rows 9–24 of Table 10. Control by each of the four non-relational elements of graphs and descriptions was established by the use of four different incorrect comparison stimuli for each type of interaction. Because these incorrect comparisons contained accurate references to all of the relational elements, and inaccurate references to one of the four non-relational elements, they are referred to as Co + −. For example, to establish control by IV1 of graphs and descriptions of each of the four types of interaction, a graph sample stimulus was presented with its Co++ and a Co − + that contained an inaccurate reference to IV1. Because a participant could only distinguish the Co++ from Co + − by attending to one of the four non-relational elements, these trials served to establish control by the four non-relational elements.

Once participants mastered non-relational training, decontextualized targeted training began, the goal of which was to insure control by Interact:d in the context of different sets of non-relational elements. Trials used during this training are represented by rows 25–40 of Table 10. During these trials, a graph sample was presented with three comparison stimuli: the Co++, Co + −, and Id-. Because a participant

Table 10 A symbolic representation of the 'A' training protocol. Numbers in the "Interaction type" column refer to the graphs shown in Fig. 1; numbers in the "Comparisons" columns refer to the descriptions shown in Fig. 1 and Table 3 and Tables 6, 7, 8, and 9

Source of control	Row #	NO	XOVR	DVR	SYN	# Cos	+ +	− −	Id-	Ip-	Is-	IV1	IV2	Dir Lo	Dir Hi
Graph:Trad.C.D	1	1				2	5	6							
	2		2			2	6	5							
	3			3		2	7	8							
	4				4	2	8	7							
Interact:d	5	1				2	5		9						
	6		2			2	6		10						
	7			3		2	7		11						
	8				4	2	8		12						
IV1	9	1				2	5					21			
	10		2			2	6					25			
	11			3		2	7					29			
	12				4	2	8					33			
IV2	13	1				2	5						22		
	14		2			2	6						26		
	15			3		2	7						30		
	16				4	2	8						34		
Dir:Lo	17	1				2	5							23	
	18		2			2	6							27	
	19			3		2	7							31	
	20				4	2	8							35	
Dir Hi	21	1				2	5								24
	22		2			2	6								28
	23			3		2	7								32
	24				4	2	8								36
Interact:d & IV1	25	1				3	5		9			21			
	26		2			3	6		10			25			
	27			3		3	7		11			29			
	28				4	3	8		12			33			
Interact:d & IV2	29	1				3	5		9				22		
	30		2			3	6		10				26		
	31			3		3	7		11				30		
	32				4	3	8		12				34		
Interact:d & Dir:Lo	33	1				3	5		9					23	
	34		2			3	6		10					27	
	35			3		3	7		11					31	
	36				4	3	8		12					35	
Interact:d & Dir Hi	37	1				3	5		9						24
	38		2			3	6		10						28
	39			3		3	7		11						32
	40				4	3	8		12						36

could distinguish the Co+ + from Co + − and Id- by attending to Interact:d and all four non-relational elements, such trials insured control by Interact:d and all four non-relational elements. Conversely, selection of one of the incorrect comparison stimuli would indicate a lack of control by either Interact:d or one of the non-relational elements. For instance,

Table 11 A symbolic representation of the 'B' training protocol. Numbers in the "Interaction type" column refer to the graphs shown in Fig. 1; numbers in the "Comparisons" columns refer to the descriptions shown in Fig. 1 and Tables 3, 4, and 5

Source of control	Row #	NO	XOVR	DVR	SYN	# Cos	+ +	− −	Id-	Ip-	Is-	IV1	IV2	Dir Lo	Dir Hi
Interact:d only	1	1				2	5		9						
	2		2			2	6		10						
	3			3		2	7		11						
	4				4	2	8		12						
Interact:p only	5	1				2	5			13					
	6		2			2	6			14					
	7			3		2	7			15					
	8				4	2	8			16					
Intersect only	9	1				2	5				17				
	10		2			2	6				18				
	11			3		2	7				19				
	12				4	2	8				20				
Interact:d, Interact:p, & Intersect	13	1				3	5		9	13	17				
	14		2			3	6		10	14	18				
	15			3		3	7		11	15	19				
	16				4	3	8		12	16	20				

if a participant was attending to Interact:d, but not IV1, the participant would select the two comparisons that contained accurate references to Interact:d—the Co++, which contained accurate references to Interact:d and IV1, and Co − +, which contains an accurate reference to Interact:d and an inaccurate reference to IV1, equally frequently. Similarly, if a participants was attending to IV1 but not Interact:d, the participant would select the two comparisons that contain accurate references to IV1—the Co++, which contains accurate references to IV1 and Interact:d, and Id-, which contains an accurate reference to IV1, but an inaccurate references to Interact:d, equally frequently. Thus, a lack of control by Interact:d or one of the four non-relational elements would result in selection of the incorrect comparison stimuli. These trials, then, assured maintenance of previously established control by Interact:d and the four non-relational elements.

Participants were required to respond with 100 % accuracy in A-training that included a test block in which all trials were presented without informative feedback. Thus, maintenance of control by Interact:d and all four non-relational elements in the absence of feedback was assessed prior to the writing test. The A-training protocol was followed by the administration of the writing test.

Training to establish the 'B' stimulus control relations ('B' training procedure) 'B' training used the same logic described in A training to establish the forms of targeted

conditional control listed in the B column of Table 2. Specifically, B training was designed to establish control by the three relational elements, Interact:d, Interact:p, and Intersect. For this reason, B training featured the use of the three Co − + stimuli. One was the Id-, which contained an inaccurate reference to Interact:d. The others contained inaccurate references to Interact:p and Intersect, and are referred to as Ip- and Is-, respectively. The sources of control established, and the sample and comparison stimuli used to establish these sources of control, are elaborated in Table 11.

Rows 1–4 of Table 11 represent the trials presented to establish control by Interact:d for all four types of interaction. During these trials, a graph sample was presented with its Co+ + and one Co − +, which was the Id-. Because the Id-contained an inaccurate reference to Interact:d, a participant could only distinguish the Co++ from Id- by attending to Interact:d. Thus, consistent correct responding on such trials would require targeted conditional control by Interact:d. This same logic was used to establish control by Interact:p, using trials represented by rows 5–8, and Intersect, using trials represented by rows 9–12.

Following the independent and sequential establishment of control by Interact:d, Interact:p, and Intersect, maintenance of control by all three relational elements was assured by trials represent by rows 13–16. These trials featured the presentation of graph samples with four comparison stimuli: the Co+ +, Id-, Ip-, and Is-. Once again, a lack of control by Interact:d,

Interact:p, or Intersect would result in selection of one of the incorrect comparison stimuli. For instance, participants attending only to one relational element, such as Interact:d would select the comparisons containing accurate references to Interact:d—the Co++, Ip-, and Is-, with equal frequency. Similarly, participants attending to only two relational elements, such as Interact:p and Intersect, would select the comparisons containing accurate references to these elements—the Co++ and Id-, with equal frequency. Consistent correct responding on such trials, then, assured maintenance of control by all three relational elements: Interact:d, Interact:p, and Intersect.

B training was followed by administration of the writing test. Participants were required to respond with 100 % accuracy in B training, which included a final test block in which all trials were presented without informative feedback. Thus, maintenance of control by Interact:d, Interact:p, and Intersect in the absence of feedback was assessed prior to the writing test. The B-training protocol finished by the administration of the writing test.

Writing test accuracy The effect of each intervention was evaluated by the accuracy of the writing test administered at the end of each intervention. Each of the three writing tests was evaluated separately, using the rubric shown in Fig. 2. One point was earned for each term if it was included in the description of a given graph. A zero was assigned if a term was not included in the written description.

Inter-observer agreement All of the written descriptions in all three tests in the experiment were assessed by three independent raters using a trial-by-trial IOA (Cooper et al. 2007). Agreement was defined as the awarding of the same number of points for a given element (either 1 or 0), using the rubric shown in Fig. 2. IOA was calculated by dividing the number of agreements by the number of agreements + disagreements, and multiplying that result by 100.

Results

Attrition

Of the 52 students who began the experiment, 43 (83 %) completed it. Of the nine who did not complete the experiment, four did not acquire the conditional discriminations in session 1 (two each assigned to the x-A-A and x-A-B groups), and five acquired all of the conditional discriminations in session 1 but did not return for session 2 (one each in the x-x-x, x-A-A, and x-A-B groups, and two in the x-x-B group). As a result, the data reported below were collected from 12 participants in the x-x-x group, 10 in the x-AA group, 10 in the x-A-B group, and 11 in the x-x-B group.

### Acquisition of conditional discriminations

The average number of trials needed to acquire one conditional discrimination in each group is shown in Fig. 3, along with the scheduled minima for each condition. All groups acquired the conditional discriminations rather rapidly and in somewhat more trials than the scheduled minima. As noted in the Procedure, the A condition involved the establishment of conditional control by the graphs, targeted conditional control by Interact:d and the non-relational elements, and decontextualized conditional control by Interact:d in combination with the non-relational elements. Since there were no apparent differences in the number of trials to acquisition for each of these types of conditional relations, their data were aggregated and treated together. The B condition involved the establishment of the targeted conditional discriminations Interact:d, Interact:p, and Intersect, each in the context of one of the four non-relational elements (IV1, IV2, Dir Lo, and Dir Hi). Since there were no apparent differences in the number of trials to acquisition, these data were aggregated and treated together. There were, however, some small but significant differences found within and across three of the experimental groups.

The x-A-A group required significantly more trials than the scheduled minimum to acquire the discriminations during initial A training, as confirmed by a one-sample t test, $t(9) = 2.3$, $p < .05$, Cohen's $d = .7$, and during repetition of A training, the x-A-A group required a similar number of trials to re-acquire the trained relations, as confirmed by a paired samples t test, $t(9) = 1.76$, $p > .05$, Cohen's $d = .6$. Thus, the initial acquisition of the A-based conditional relations did not shorten the re-acquisition of the same relations. Participants in the x-x-B group also required significantly more trials than the scheduled minimum to learn the B conditional discriminations, as confirmed by a one-sample t test, $t(10) = 3.5$, $p < .01$, Cohen's $d = 1.1$. Last, the participants in the x-A-B group also required significantly more trials than the scheduled minimum to learn the conditional discriminations during initial A training, as confirmed by a one-sample t test, $t(9) = 2.7$, $p < .05$, Cohen's $d = .9$. In addition, significantly fewer trials were needed to acquire the conditional discriminations during B training than during the previously administered A training, as confirmed by a paired-samples t test, $t(9) = 2.4$, $p < .05$, Cohen's $d = .8$. Finally, significantly fewer trials were needed to acquire the conditional discriminations during B training in the x-A-B condition than the equivalent training in the x-x-B group, as confirmed by an unpaired t test, $t(19) = 2.8$, $p < .01$, $r^2 = .3$.

Writing performances

Inter-observer agreement All written descriptions were evaluated for accuracy for sentences that described interactions, intersections, and directionality. Inter-observer-agreements

NO Interaction
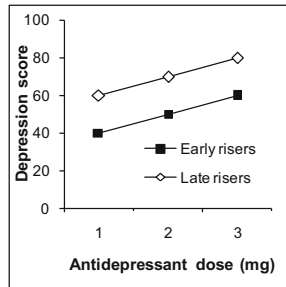
Dir Lo (4 pts): **Depression score** (1) was a **direct function** (1) of **antidepressant dosage** (1) for **late risers** (1).

Dir Hi (4 pts): In addition, **depression score** (1) was a **direct function** (1) of **antidepressant dosage** (1) for **early risers** (1).

Interact (6): **Increasing the antidepressant dosage** (1) **produced** (1) a **constant difference** (1) in **depression scores** (1) for **late risers** (1) **relative to early risers** (1).

Intersect (5): **Depression score** (1) functions for **late** (1) - and **early-risers** (1) **did not intersect** (1) at any level of **antidepressant dosage** (1).

CROSSOVER
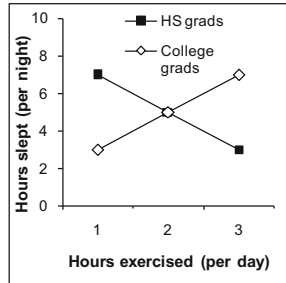
Dir Lo (4 pts): For **college graduates** (1), the number of **hours slept** (1) was a **direct function** (1) of the number of **hours exercised** (1) per day.

Dir Hi (4 pts): For **high school graduates** (1), the number of **hours slept** (1) was an **inverse function** (1) of the number of **hours exercised** (1) per day**.**

Interact (4 pts): For different numbers of hours spent exercising, then, **education** (1) **reversed** (1) the effect of **time spent exercising** (1) on the number of **hours slept** (1).

Intersect (5 pts): The **hours slept** (1) functions for individuals of **different education levels** (1) **intersected** (1) at an intermediate level of **hours spent exercising** (1).

DIVERGENT
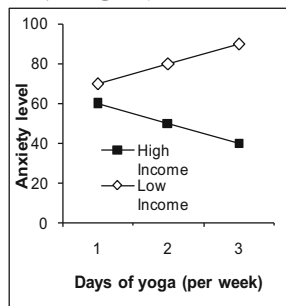
Dir Lo (4 pts): The **anxiety level** (1) was a **direct function** (1) of **days of yoga** (1) per week for **low-income individuals** (1).

Dir Hi (4 pts): The **anxiety level** (1) was an **inverse function** (1) of **days of yoga** (1) per week for **high-income individuals** (1).

Interact (4 pts): **Income level** (1) **reversed** (1) the directional effect of **days of yoga** (1) on **anxiety level** (1).

Intersect (5 pts): Finally, the **anxiety level** (1) functions for **high-** (1) and **low-income individuals** (1) **did not intersect** (1) at any number of **days of yoga** (1).

SYNERGISTIC
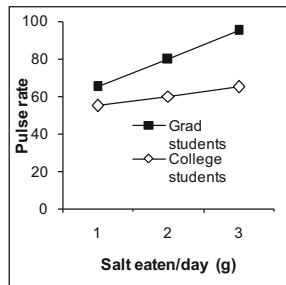
Dir Lo (4 pts): In addition, **pulse rate** (1) was a **direct function** (1) of **salt consumption** (1) for **college students** (1).

Dir Hi (4 pts): **Pulse rate** (1) was a **direct function** (1) of **salt consumption** (1) for **graduate students** (1).

Interact (6 pts): Increasing the amount of **salt eaten** (1**) produced** (1) **a growing difference** (1) in **pulse rate** (1) for **graduate students** (1) relative to **college students** (1).

Intersect (5 pts): **Pulse rate** (1) functions for **graduate and college students** (1) **did not intersect** (1) at any level of **salt consumption** (1).

Fig. 2 The grading rubric used to evaluate participants' written descriptions. Points are earned for the presence of bold font terms. The numbers of points possible for an entirely correct written production of each sentence are shown in parentheses. Writing accuracy scores were calculated by totaling the number of points earned for entire descriptions and for each type of sentence

varied from 92 % to 99 % for each type of sentence across all three writing tests. These results documented the reliability of the ratings of accuracy assigned to all of the writing samples produced by the participants in the experiment.

Intervention effects on yields The top panel of Fig. 4 displays the percentage of participants in each group who

achieved two different accuracy levels (75 % or 90 %) on the final writing test. These values correspond to letter grades of C+ and A, respectively. In the x-x-x or x-A-A groups 0 % of the participants wrote descriptions that were at least 75 % accurate on the final writing test. However, in the x-x-B group, 45 % of participants wrote descriptions that were at least 75 % accurate, suggesting that B training alone produced an
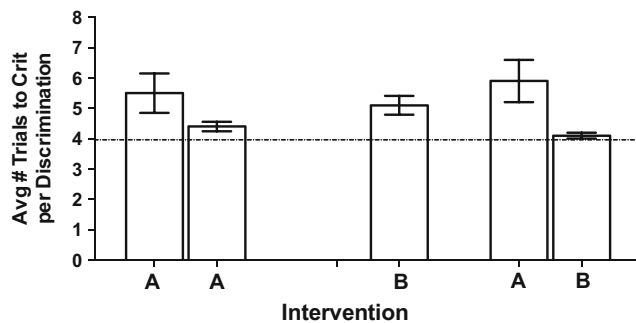
Fig. 3 The average number of trials required to reach mastery criterion by participants in the three groups: x-A-A (left-most set of bars), x-x-B (middle bar) and x-A-B (right-most set of bars). The dashed line indicates the minimum number of trials needed to reach mastery

Fig. 4 The percentage of participants in each group that achieved each of two different constant performance criteria on the final writing test (**top panel**), and the average accuracy of each group on the final writing test (**bottom panel**). Error bars represent the standard error of the mean

intermediate level of writing accuracy. This hypothesis is further supported by the performance of the x-A-B group, in which the combination of A and B training resulted in 75 % accuracy scores by 90 % of participants. While the combination of A and B training enhanced the number of participants that achieved 75 % accuracy, B training alone was sufficient to produce an improvement in writing to the level of an academic grade of C+.

On the other hand, no participants in either the x-x-x, x-A-A, or x-x-B groups wrote descriptions that were at least 90 % accurate. Only the combination of A and B training, in the x-A-B group, produced writing of this accuracy level, which was achieved by 80 % of participants. Thus, only the combination of A and B training resulted in writing accuracy that reached the level of an academic grade of A.

Overall writing accuracy The results of the experiment can also be assessed by computing the overall accuracy of participants' descriptions at the end of all training. Such a measure would reflect the accuracy for all of the relational and non-relational elements in the interaction, intersection, and directionality statements. These data are presented in the bottom panel of Fig. 4. An independent groups ANOVA found significant a significant difference between the overall accuracy scores of the four groups, $F(3, 39) = 79.06$, $p < .01$. The following significant differences between the four groups were found by a Tukey's post-hoc test.

In the last of the writing tests in the x-x-x- condition, the participants wrote relatively inaccurate and incomplete descriptions of the information conveyed in the interaction-based graphs. After A-training alone, accuracy of the written descriptions increased to an intermediate value that was significantly greater than that observed after no training (Tukey's LSD $p < .01$). After B training alone, the participants produced an intermediate accuracy of the written descriptions that was significantly greater than that observed after no training (Tukey's LSD $p < .01$). Accuracy of the written descriptions
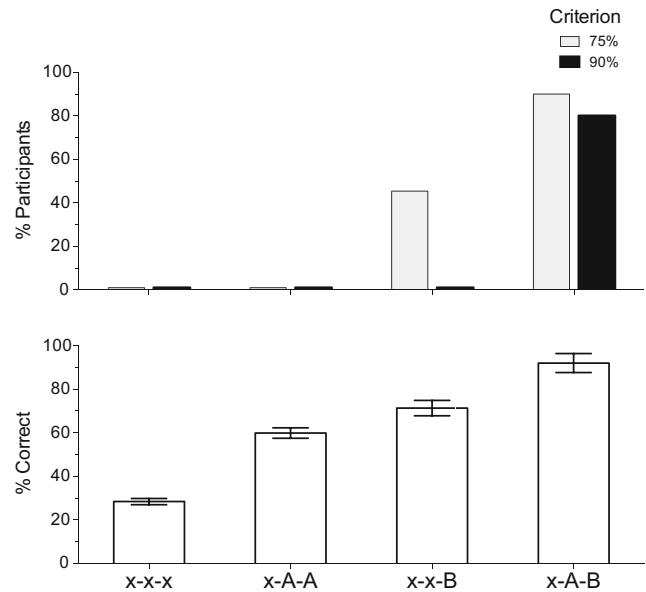
after B-training alone was also equal that observed after A-training alone.

When A training was followed by B training (in the x-A-B condition), the written descriptions were almost completely accurate; only one participant produced descriptions less than 84 % accurate. When compared to the effects of test repetition alone, this increment in writing accuracy was significant (Tukey's LSD $p < .01$). Accuracy after A + B training was also significantly greater than that observed after A-training alone (Tukey's LSD $p < .01$) and B-training alone (Tukey's LSD $p < .01$). Indeed, the sum of A training alone and B training alone approximated the accuracy obtained when A and B training were used together. Thus, the effects of A-training and B-training worked in an additive manner to produce accurate written descriptions of the interactive effects of two variables on behavior.

Writing accuracy by sentence type The written descriptions of each graph contained sentences that characterize the interaction of the two independent variables (Interact), the intersection of the two functions (Intersect), and directionalities of the functions in the graphs (Direction). Thus, the effects shown in the bottom panel of Fig. 4 could reflect the differential effects of A- and B-training on the production of any or all of the three types of sentences. Thus, the effects of each training condition on writing accuracy were analyzed separately for the sentences that described the interactions, the intersection of the functions of the interaction-based graphs, and the directionalities of the functions on the graphs. Each analysis is presented on a separate panel in Fig. 5.
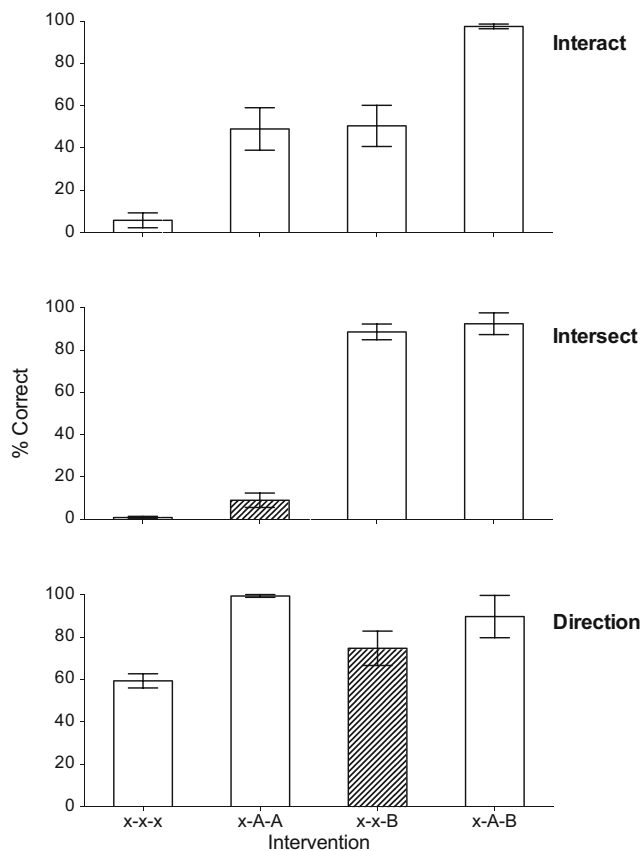
Fig. 5 The average accuracy of each of the three sentences produced by each group. Shaded bars indicate performances that reflect generalization, and error bars represent the standard error of the mean

Interaction statements With respect to the accuracy of the interaction statement, an independent groups ANOVA found a significant difference between the accuracy scores of the four groups, $F(3, 39) = 28.06$, $p < .01$. Test repetition resulted in a very low level of accuracy. A-training alone and B-training alone each produced the same intermediate levels of accuracy, both of which were significantly different from that measured after test repetition (Tukey's LSD ps < .001). Finally, training with the combination of A and B produced written descriptions that were almost completely accurate, and significantly greater than that observe after no training (Tukey's LSD $p <$ .001), A-training alone (Tukey's LSD $p < .01$) and B-training alone (Tukey's LSD $p < .001$). The accuracy under the x-A-B condition was equal to the sum of the accuracies produced by A-training alone and B-training alone. Thus, the emergence of accurate written descriptions of the interactive effects of two variables on behavior was produced by the additive effects of the prior establishment of conditional control during A-training in combination with B-training.

Intersection statements With respect to the accuracy of the intersection statements, an independent groups ANOVA found a significant difference between the accuracy scores

of the four groups, $F(3, 39) = 209.6$, $p < .01$. Test repetition resulted in a minimal level of accuracy. A training alone resulted in a low level of writing accuracy. In contrast, B training alone produced a large and significant increase in writing accuracy nearly as great as that produced by B-training in combination with A-training (Tukey's LSD ps < .01). Thus, the increment in the writing of accurate descriptions of intersection was based predominately on B-training alone.

Because the x-A-A condition did not involve the training of control by the intersection elements in the graphs and the printed descriptions, the slight, non-significant increment in written performance after A-training reflects generalization of the effects of training. Because B training established targeted conditional control by the intersection statements, B training was responsible for the emergence of very accurate written descriptions of the directional effects of two variables on behavior.

Directionality statements With respect to the directionality statements, an independent groups ANOVA found a significant difference between the accuracy scores of the four groups, $F(3, 39) = 7.43$, $p < . 01$. Test repetition resulted in an intermediate level of accuracy. This indicated the presence of stimulus control by the elements of the directionality relations prior to any experimenter-based training. The A-training condition alone produced nearly maximal levels of written accuracy, significantly more than retesting (Tukey's LSD $p <$ .01). By contrast, B-training alone produced a small, non-signficiant increment in accuracy when measured against test repetition alone. Finally, after the combination of A- and B-training, the accuracy of descriptions of the directionalities of the function in the interaction-based graphs was almost the same as that observed after A training alone (Tukey's LSD p < .01). Therefore, the emergence of accurate written descriptions of the intersections of the functions in the graphs was produced by A-training alone. This occurred because A-training explicitly established conditional control by the directionalities of the functions in the graphs and the printed descriptions of the directionalities of these functions. The B-training condition did not involve the explicit establishment of relations between the directional elements in the graphs and printed descriptions. Therefore, the increment in performance after B-training alone reflected generalization.

Summary The effects of A and B training alone and together had differential effects on the accuracies of written descriptions of the interactive effects of two variables on behavior, on the intersection of functions in interaction based graphs, and on the directionalities of the functions in interaction based graphs. Although neither A-nor B-training alone produced complete accuracy of written descriptions, their combination was responsible for the induction of such a performance.

Intervention effects of A- and B-training within and between conditions

The analyses presented above focused on the final effects of each three-stage intervention package, and was measured using the last of the written descriptions of the interaction-based graphs. In the experiment, however, the A and B training protocols were introduced serially, with A preceding B, and writing tests administered in after each protocol. Thus, it is also possible to isolate the effects of each protocol by considering writing accuracy after the administration of each. These effects are seen in the panels in the right column in Fig. 6.

When the overall accuracy of the written descriptions was considered, a low level of accuracy occurred after not training, a higher level of accuracy occurred after A training, and a high level of accuracy occurred after B training. The same pattern of writing accuracy occurred in the sentences that described

the interaction of the two variables on behavior. When the accuracy of the written descriptions of the intersection sentences were considered, very low and high levels of accuracy occurred after A- and B-training, respectively. When the accuracy of the written descriptions of the directionalities of the functions in the graphs were considered, similarly high levels of accuracy occurred after A- and B-training. Thus, these within group comparisons of the effects of A and B training on writing descriptions of interactions were the same as those drawn from the between group analyses presented in Figs. 4 and 5.

Intervention regularity within and across conditions

The same intervention was used a different number of times across some of the conditions. To what extent did repetition of the same intervention influence the accuracy of the written
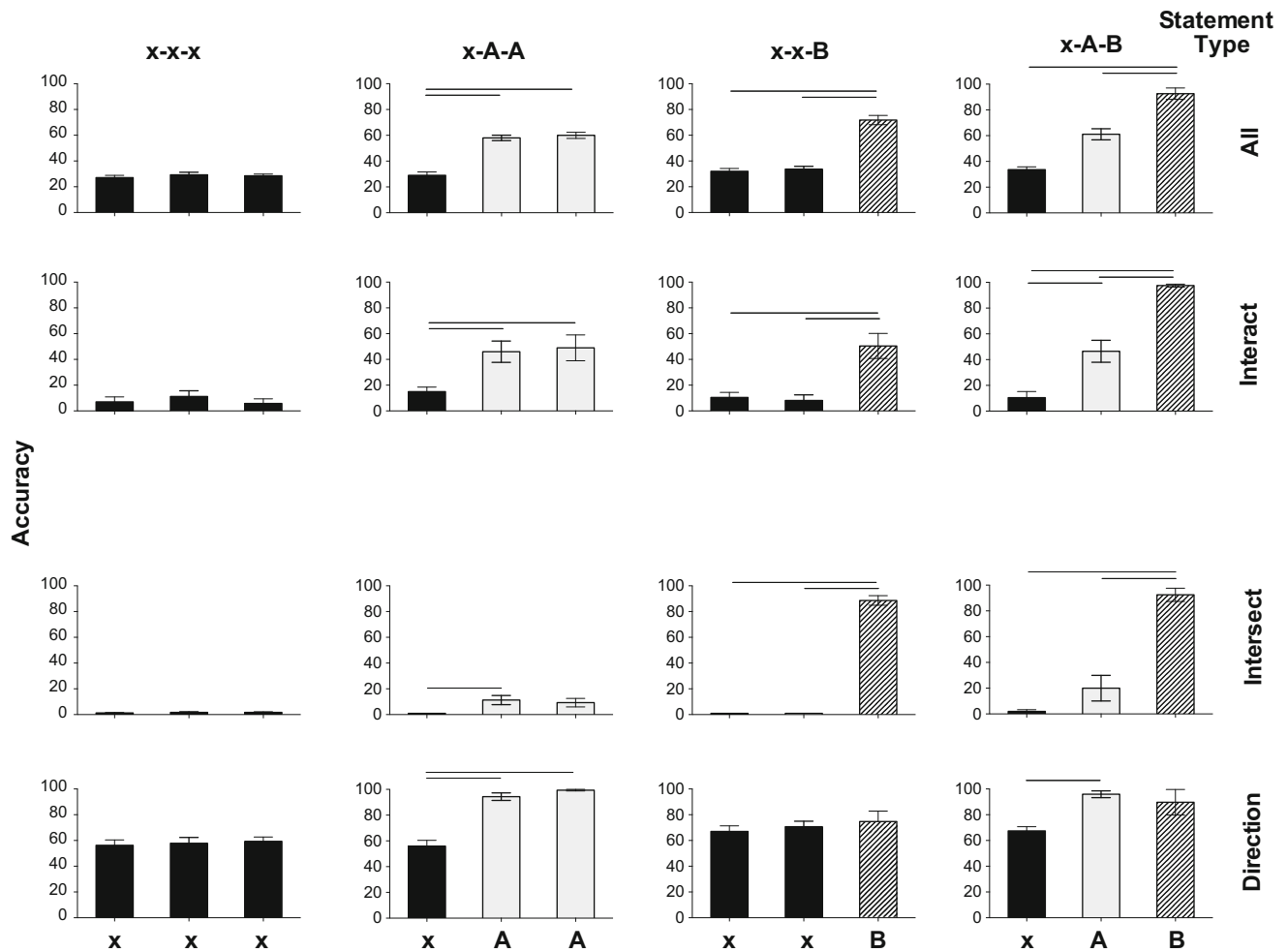

Fig. 6 Accuracy of each of the three statements following each intervention. Black bars indicate performances following no intervention; gray bars indicate performances following A training, and striped bars indicate performances following B training. Error bars represent the standard error of the mean. Endpoints of the horizontal bars indicate significant differences found by a post-hoc test after a repeated-measures ANOVA

descriptions? Results are shown in Fig. 6, which plots the accuracy of the written descriptions produced after each intervention (see the bars in each panel) in each condition (see the panels in each column) for all and each type of sentence (see each row). Note that the right-most bar in each panel is the same at that depicted in Fig. 5.

No training (x) No intervention (i.e., x) was used three times in the x-x-x condition, two times in the x-x-B condition, and one time in the x-A-A and x-A-B conditions. Repeated testing did not produce a change in overall writing accuracy (top panel). When the Interaction sentences were considered (second row, gray bars), very similar quite low levels of accuracy were produced following each no-training intervention (x). A similar result was obtained for the accuracies in the intersection sentence (third row, gray bars), and also with the directionality sentences (bottom row, gray bars). The levels of accuracy varied, however, across sentence type. Very low accuracy was found for the interaction sentences, even lower for the intersection sentences, and intermediate for the directionality sentences. To summarize, writing accuracy was not influenced by test repetition, regardless of sentence content.

A-training The training of graph based discriminations followed by targeted Interact:d training and decontextualized non-relational training (A) occurred twice in the x-A-A condition and once in the x-A-B condition. To what extent did the initial level of A training produce similar effects? Overall (top panel), A training produced a significant improvement in writing accuracy intermediate level of writing accuracy. When the Interaction sentences are considered (second row, black bars), very similar levels of accuracy were produced after the completion of the first administration of A-training. A similar result was obtained for the accuracies in the Intersection sentence (third row, black bars) and the Directionality sentences (bottom row, black bars). In addition, the repetition of A-training did not change the accuracy of the descriptions for each of the sentences. Finally, the levels of accuracy varied with sentence type: intermediate for the interaction sentences, very low for the intersection sentences, and high for the directionality sentences. To summarize, A-training produced similar effects on writing accuracy across experimental conditions, and these improvements were greatest in descriptions of the Interaction and Directionality statements. Control by both of these elements was explicitly established by A training. Further, the magnitude of these improvements did not increase with repetition of A training.

The effects of A-training on B-training

As noted above, each condition involved the sequential administration a few different intervention packages: no training

or x, A, and B. Thus, it is also possible to determine whether the accuracy of the descriptions written after B training was influenced by prior training. Specifically, B training in the x-x-B condition was not preceded by any other form of training. In contrast, B training in the x-A-B condition was preceded by A-training. The accuracies of the descriptions written after B training in these two conditions would indicate how A- and B-training influenced the accuracy of the last descriptions written by the students. Further, since this analysis can be done for each type of sentence, the effect on each sentence can also be isolated.

In the x-x-B condition, B training produced an intermediate level of overall writing accuracy (top row, striped bars). The improvement in writing accuracy was greatest in descriptions of the Interaction sentences (second row, striped bars), and the Intersection sentences (third row, striped bars). On the other hand, B training did not significantly change the accuracy of descriptions of the Directionality sentences (bottom panel).

In the x-x-B condition, B training was not preceded by any other training. In contrast, a very high level of overall accuracy (top panel) was produced in the x-A-B condition when B-training was preceded by A-training. These results indicate that the performances in the writing test conducted after A + B training reflected the additive effects of A- training and B-training. Specifically, A-training established Interact:d relations directly, while B-training established Interact:p relations directly. Therefore both forms of training established the types of relations that provided models for the writing of accurate descriptions of interaction sentences in the writing test that followed B-training in the x-A-B condition.

When the Interaction and Intersection sentences are considered (second and third rows, striped bars), the combination of A and B training produced very high and essentially equal levels of accuracy. However, the accuracy of descriptions of the Intersection sentences in the x-A-B condition is essentially equal to those observed after B training alone.

When the Directionality sentences are considered (bottom row, striped bars), B training produced a moderately high level of accuracy in the x-x-B condition, where B-training was not preceded by any other form of training, and a similar level of accuracy in the x-A-B condition, where B training was preceded by A training.

Discussion

Synopsis

It is very difficult for students majoring in psychology to write accurate and complete descriptions of the interactive effects of two variables on behavior. In the present experiment, most students wrote accurate and complete paragraph-length written descriptions of such graphs following the establishment of

traditional, targeted, and decontextualized conditional discriminations between key elements in the graphs and the corresponding elements of their printed descriptions. Further, the writing of accurate descriptions emerged without any direct training of the writing responses. The graph-text conditional discriminations in their various forms were established using selection-based responding. Writing, on the other hand, was an expressive or production-based repertoire. Therefore, the establishment of a selection-based repertoire resulted in the immediate emergence of a rather extensive production-based repertoire.

The results of the present experiment provided an efficient automated basis that suggests a solution to the problem stated above.

Effects of components of A and B training alone and together

Effects of B-training alone The production of complete and accurate descriptions of the interactions and intersections had to reflect at least three stimulus control topographies: targeted conditional control by phrases describing the interaction (Interact:d and Interact:p) and the phrase describing the intersection of the functions in the graphs (Intersect). These had to be present for each of the four graphs, for a total of four "sets" of three stimulus control topographies. When these topographies were established in the context of only one set of non-relational elements, as in the B training protocol, the participants wrote descriptions that were about 50 % accurate. Therefore, establishing targeted conditional control by graph-Interact:d, graph-Interact:p, and graph-Intersect relations together was not sufficient to induce the writing of accurate descriptions of the interaction-based graphs.

Effects of A-training alone A training involved the establishment of conditional relations between the graph and any element of printed text (traditional training), targeted conditional relations between graphs and one set of phrases that described the interactive effects of the variables on behavior (the Interact:d relations), and decontextualized stimulus control by the Interact:d relations in combination with four non-relational elements. To what extent did each of these components induce accurate writing of descriptions of the interactions?

Part of the A training protocol involved the use of traditional conditional discrimination training to establish a discrimination among the various graphs used as sample stimuli and conditional relations between each graph and one unspecified element of the printed paragraphs used as comparison stimuli. While such training was not studied in isolation in the present experiment, it was used in a virtual replication of the present experiment by Spear and Fields (submitted), and produced very inaccurate written descriptions. In the present

experiment it is reasonable to assume that the traditional conditional discrimination training used in the A-training component did not influence the emergence of accurate written descriptions of interactions. This finding is also consistent with the failures to describe medical diagnoses and features of research designs by the participants in the experiments reported by Walker et al. (2010) and Walker and Rehfeldt (2012).

The B training protocol established control by each of the relational elements in the context of only one set of non-relational elements. It did not establish decontextualized control by the combination of relational elements, such as Interact:d, Interact:p and Intersect, and the non-relational elements of graphs and descriptions that would also be featured in accurate descriptions of the interactions. Thus, it is possible that the establishment of control by the relational elements in the context of more than one set of non-relational elements would have produced accurate written descriptions of the interactions. This issue was addressed in part with the A training protocol that established control by the Interact:d relational elements in the context of four sets of non-relational elements. This training also resulted in the production of descriptions that were about 50 % accurate. Thus, the establishment of decontextualized conditional control by one of the elements of the interaction (Interact:d) in context of many sets of non-relational elements was partially responsible for the induction of accurate written descriptions.

Effects of A and B training together The very high level of writing accuracy that emerged only after the establishment of the conditional relations established during A and B training consisted of improvements in the production of descriptions of all of the features of the graphs: the directionalities of the functions, the interaction of the independent variables, the intersection of the functions, and the specification of the non-relational variables. These results demonstrate that only the combination of A and B training established all sources of stimulus control necessary to produce nearly complete and accurate descriptions of the interactions.

Sources of stimulus control after A and B training The exact sources of stimulus control established by the A and B protocols are shown in Table 2. There are a total of 16 forms of control that can be established for each of the four types of interaction, for a total of 64 forms of conditional stimulus control. Theoretically, all of them would have to be intact for an individual to produce a complete and accurate description of a graph. Traditional conditional discrimination training, as reported in an earlier study (Spear & Fields, submitted), established one of these sources of control for each of the four interaction types (graph and any correlated element of the printed description of the graph) for a total of four of the 64 forms of control. This procedure resulted in the writing of

descriptions that were approximately 20 % accurate. Similar results by Walker et al. (2010) and Walker and Rehfeldt (2012).

The A training protocol in the present experiment established six conditional stimulus control topographies for each of the four types of interaction, for a total of 24 of the 64 forms of control. These 24 forms of control included targeted conditional stimulus control for one graph-phrase relation (Interact:d) in the context of four non-relational elements. This training resulted in the writing of descriptions that were approximately 70 % accurate. The addition of B training, which established two additional forms of targeted conditional control for each of the four types of interaction, or a total of eight forms of control, results in the writing of near completely accurate descriptions. Thus, establishing control by 32 of the 64 possible forms of control, as accomplished by A and B training together, resulted in nearly complete writing accuracy. These performances occurred only when the training routines were designed to establish three targeted graph-phrase conditional relations (Interact:d, Interact:p, and Intersect) and control by one of the three already established graph-phrase relations (Interact:d) in the context of the four non-relational elements that characterized the interactive effects of two variables on behavior.

It is possible that the 32 forms of conditional and control established by the current procedures were not necessary to induce the production of accurate descriptions. Rather, the establishment of a different set of 32 forms of control—including, for instance, Intersect in the context of the four non-relational elements—might suffice. It is also possible that the establishment of an as-yet-unknown smaller number of forms of control would result in the production of accurate descriptions. Both of these issues require further research. The results of such a research endeavor might then identify the necessary and sufficient conditions needed to induce the writing of accuracy descriptions of the interactive effects of two variables on behavior.

Mechanisms of effect: Emergent repertoires and function transfer

Emergent repertoires As stated previously, the production of complete and accurate descriptions of necessity implies the presence of all 64 forms of targeted and decontextualized conditional stimulus control. The combination of A and B training established only half of these, and yet resulted in the writing of near completely accurate descriptions. The fact that this outcome was achieved indicates generalization of writing performance from trained to untrained forms of control. For instance, control by IV1 was established by A training, and control by Intersect was established by B training. Thus, control by each of these elements was established in isolation. After the combination of A and B training, however, these

two elements exerted control in combination, facilitating production of an accurate description of the intersection that included accurate references to both Intersect and IV1. The emergence of such an accurate description is an example of contingency adduction (Andronis et al. 1997) or recombinative generalization (Goldstein 1983; Goldstein et al. 1987), as these two elements had not been directly linked during prior conditional discrimination training but, nonetheless, together exerted control over writing behavior. Such emergence of control by untrained combinations of elements of complex stimuli has also been found in other studies but the stimuli were simpler than those used in the present experiment (Alonso-Alvarez and Perez-Gonzalez 2006; Alonso-Alvarez and Perez-Gonzalez 2011; Perez-Gonzalez and Alonso-Alvarez 2008).

Function transfer Recombinative generalization, however, does not explain the main finding of the current research: a generalized improvement in writing, a production-based response, following selection-based training. A possible explanation for this outcome is based on transfer of function: participants entered the experiment with a text-copying repertoire (i.e., the ability to produce a complete and accurate copy of text presented to them), and the training procedures used in the current study likely facilitated the transfer of this text-copying function from the printed paragraphs to the graphs. In combination with recombinative generalization, the current selection-based procedures could produce a generalized improvement in writing behavior. If this account is accurate, failure of the current training procedures to produce an improvement in writing could result from a lack of either function transfer or generalization, and could be treated by conducting additional training to establish the some or all of the additional 32 forms of control not established by the current procedures.

Most studies of function transfer have been conducted in the context of equivalence classes: some function is acquired by one class member and then generalizes with a constant high probability to the other class members. In the present experiment, equivalence class formation was not used to study function transfer. Rather, a set of graph-text conditional relations were established, after which the behaviors presumed to occur in the presence of one of the stimuli generalized to the other member of the conditional relation. In addition, the stimuli in the visual-visual conditional relations were quite complex and the response that generalized involved the production of paragraph-length written descriptions. Thus, function transfer did not depend on the prior formation of equivalence classes, but rather of conditional discriminations.

It would be of interest, however, to expand the behavioral preparation by establishing a second type of conditional discrimination, for instance, between the name of the type of interaction and the graph of that type of interaction.

Thereafter, two important questions arise: first, would the presentation of any one of the stimuli representing one of the types of interaction evoke the selection of the other stimuli representing that interaction, i.e., would equivalence classes form? Second, would the presentation of any one of the stimuli, such as the name of the interaction, evoke the writing of a complete and accurate description of the graph depicting that type of interaction, i.e., would the writing function transfer to all members of the same class?

Summary and conclusions

In the current experiment, conditional discrimination training procedures that established traditional, targeted, and decontextualized conditional control by many elements of complex stimuli were found to produce an improvement in writing behavior. This research replicates and remedies the deficits of traditional conditional discrimination training reported by others (Walker et al. 2010; Walker and Rehfeldt 2012). Further, the improvement in writing behavior is reflective of recombinative generalization, as stimulus control topographies established separately came to exert control over writing together, and also of transfer of function, as the procedures did not directly train writing. Whether these procedures result in the emergence of other types of stimulus relations (i.e., symmetry) and the applicability of these procedures to other stimulus materials are issues for future research.

References

Alonso-Alvarez, B., & Perez-Gonzalez, L. A. (2006). Emergence of complex conditional discriminations by joint control of compound samples. *The Psychological Record, 56,* 447–463.

Alonso-Alvarez, B., & Perez-Gonzalez, L. A. (2011). Derived control by compound and single sample stimuli in a matching-to-sample task in children. *Psicothema, 23,* 415–423.

Andronis, P. T., Layng, T. V. J., & Goldiamond, I. (1997). Contingency adduction of "symbolic aggression" by pigeons. *The Analysis of Verbal Behavior, 14,* 5–17.

Behavior Analyst Certification Board. (2012). *Fourth Edition Task List.* Retrieved from http://www.bacb.com/Downloadfiles/TaskList/Fourth_Edition_Task_List.pdf

Belanich, J., & Fields, L. (2003). Generalized equivalences class as response transfer networks. *The Psychological Record, 53,* 373–413.

Cooper, J., Heron, T., & Heward, W. (2007). *Applied behavior analysis.* New Jersey: Pearson Education.

Cowley, B. J., Green, G., & Braunling-McMorrow, D. (1992). Using stimulus equivalence procedures to teach name-face matching to adults with brain injuries. *Journal of Applied Behavior Analysis, 25,* 461–475.

Eikeseth, S., & Smith, T. (1992). The development of functional and equivalence classes in high- functioning autistic children: The role of naming. *Journal of the Experimental Analysis of Behavior, 58,* 123–133.

Fields, L., & Spear, J. (2012). Measuring joint stimulus control by complex graph/description correspondences. *The Psychological Record, 62,* 279–294.

Fields, L., Adams, B. J., Buffington, D. M., Yang, W., & Verhave, T. (1996). Response transfer between stimuli in generalized equivalence classes: A model for the establishment of natural kind and fuzzy superordinate categories. *The Psychological Record, 46,* 665–684.

Fields, L., Travis, R., Roy, D., Yadlovker, E., de Aguiar-Rocha, L., & Sturmey, P. (2009). Equivalence class formation: A method for teaching statistical interactions. *Journal of Applied Behavior Analysis, 42,* 575–593.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education, 19*(1), 44–63.

Garfield, J., & Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematical Thinking and Learning, 2,* 99–125.

Goldstein, H. (1983). Recombinative generalization: Relationships between environmental condtions and the linguistic repertoires of language learners. *Analysis and Intervention in Developmental Disabilities, 3,* 279–293.

Goldstein, H., Angelo, D., & Mousetis, L. (1987). Acquisition and extension of syntactic repertoires by severely mentally retarded youth. *Research in Developmental Disabilities, 8,* 549–574.

Groskreutz, N. C., Karsina, A., Miguel, C. F., & Groskreutz, M. P. (2010). Using complex auditory–visual samples to produce emergent relations in children with autism. *Journal of Applied Behavior Analysis, 43,* 131–136.

Mulhern, G., & Wylie, J. (2004). Changing levels of numeracy and other core mathematical skills among psychology undergraduates between 1992 and 2002. *British Journal of Psychology, 95,* 355–370.

Perez-Gonzalez, L. A., & Alonso-Alvarez, B. (2008). Common control by compound samples in conditional discriminations. *Journal of the Experimental Analysis of Behavior, 90,* 81–101.

Perez-Gonzalez, L. A., Herszlikowicz, K., & Williams, G. (2008). Stimulus relations rnalysis and the emergence of novel intraverbals. *The Psychological Record, 58,* 95–129.

Rosales, R., Rehfeldt, R. A., & Lovett, S. (2011). Effects of multiple exemplar training on the emergence of derived relations in preschool children learning a second language. *Analysis of Verbal Behavior, 27,* 61–74.

Sella, A. C., Ribeiro, D. M., & White, G. W. (2014). Effects of an online stimulus equivalence teaching procedure on research design open-ended questions performance of international graduate students. *The Psychological Record, 64,* 89–103.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research, 14,* 5–13.

Sidman, M., & Cresson, O. (1973). Reading and crossmodal transfer of stimulus equivalences in severe retardation. *American Journal of Mental Deficiency, 5,* 515–523.

Skinner, B. F. (1957). *Verbal behavior.* Englewood Cliffs: Prentice Hall.

Walker, B., & Rehfeldt, R. A. (2012). An evaluation of the stimulus equivalence paradigm to teach single-subject design to distance education students via blackboard. *Journal of Applied Behavior Analysis, 45,* 329–344.

Walker, B., Rehfeldt, R. A., & Ninness, C. (2010). Using the stimulus equivalence paradigm to teach course material in an undergraduate rehabilitation course. *Journal of Applied Behavior Analysis, 43,* 615–633.