On the structure of relational responding

Jamie Cummins[1], Maura Nevejans[1], Dylan Colbert[2], & Jan De Houwer[1]

[1]Department of Experimental-Clinical and Health Psychology, Ghent University, Belgium

[2]Department of Psychology, Dublin Business School, Ireland

Author Note

JC, MN, and JDH, Department of Experimental Clinical- and Health Psychology, Ghent University, and DC, Department of Psychology, Dublin Business School. JC and MN contributed equally. Funding for this research was provided by Ghent University Grant BOF16/MET_V/002 to JDH. Correspondence concerning this article should be sent to jamie.cummins@ugent.be.

Ethical approval was obtained through the Ghent University Ethics Committee, approval number 2020/74. Informed consent was obtained from all participants prior to completion of the experiment.

Abstract

Relational Frame Theory (RFT, S. C. Hayes et al., 2001) predicts that some topographies of relational responding should map onto one another more closely than others. By extension, training one type of relational responding should differentially improve other relational responses as a function of their relatedness to the trained relation. We investigated these predictions in two experiments. Using an elaborated version of the Relational Abilities Index (Colbert et al., 2020) in Experiment 1, we investigated the correlations between various types of relational responding. In Experiment 2, we then provided targeted relational training to two separate groups. Both groups trained on a different relation (either difference or containment relations). We found that this training not only increased performance on the directly trained relation, but also performance on other related relations.

Keywords: Relational Frame Theory, relational training, derived relational responding, relational reasoning, cognition

On The Structure of Relational Responding

The ability to reason relationally about the world is viewed by almost all psychologists as a critical feature of human language and cognition. This consensus spans across the fields of linguistics, cognitive science, computational modelling, behaviour analysis, and a host of other areas (McLoughlin et al., 2020). A variety of theoretical accounts have been forwarded which seek to explain, and make predictions about, the relationship between relational responding and cognitive performance. Relational Frame Theory (RFT, S. C. Hayes et al., 2001) provides one such account of relational responding from a behaviour-analytic perspective. In effect, RFT proposes that there are two key features which underpin human cognition; namely, the ability to *learn* relationships between stimuli that share no physical features, and the ability to *derive* relations between stimuli that have never been trained directly. This latter phenomenon is known as derived relational responding (for a detailed account, see S. C. Hayes et al., 2001).

Importantly, derived relational responding is conceived as a form of generalised operant behaviour (Healy et al., 2000; Lipkens et al., 1993). The most critical implication of this conceptualisation is that these relational responding skills (and by extension, the cognitive abilities which are underpinned by these skills) may be altered and improved through training. Although a somewhat startling claim to non-behaviourists, this idea has received tentative, but growing, support in recent years. Several studies have already demonstrated that relational responding abilities in children can be improved by training those same relational skills (Barnes-Holmes, Barnes-Holmes, & Smeets, 2004; Barnes-Holmes, Barnes-Holmes, Smeets, et al., 2004; Dunne et al., 2014; J. Hayes et al., 2016). Further in line with this idea is the fact that relational skills have been shown to correlate with various measures of intelligence (Colbert et al., 2017, 2020; Gore et al., 2010; J. Hayes et al., 2016; Kirsten & Stewart, 2021; O'Hora et al., 2005, 2008) and language abilities (J. Hayes et al., 2016; Moran et al., 2014).

Given that relational skills can be trained, and that relational skills correlate with measures of intellectual ability, then by extension training relational skills may improve intellectual abilities. Based on this idea, different researchers have started to develop training programmes with this goal in mind, such as the SMART programme (Strengthening Mental Abilities with Relational Training, Cassidy et al., 2011) and PEAK training (Promoting the Emergence of Advanced Knowledge; Dixon et al., 2022). Indeed, several studies now suggest that such relational training has substantial and beneficial impacts on a range of intellectual abilities in children, (Amd & Roche, 2018; Cassidy et al., 2011; J. Hayes & Stewart, 2016; McLoughlin et al., 2021; Vizcaíno-Torres et al., 2015), adolescents (Cassidy et al., 2016; Colbert et al., 2018; McLoughlin et al., 2020), and adults (Thirus et al., 2016). Furthermore, a recent meta-analysis found an overall moderate impact of relational training on nonverbal IQ, although also noted that there are comparably few studies which have investigated this and a relatively high risk of bias in those extant studies (May et al., 2022).

Although relational training therefore represents a promising avenue for further exploration, it is important to note that little attention has been paid to the underlying structure of relational responding skills. In general, researchers typically consider several different types of relational responding (e.g., opposition [A is opposite to B], temporal [A comes before B], and quantity relations [A is more than B], to name a few). Training protocols also tend to target these different relations under the assumption that they represent different relational response classes. However, this delineation is based exclusively on topographical features (i.e., that the relational terms themselves differ), rather than functional features. However, the fact that two relational terms are formally different does not necessarily mean that they are part of separate operant response classes. For instance, it may well be the case that “temporal” relations are simply part of the broader classes of “quantity” relations (e.g., “X comes after Y” can be functionally equivalent to “X is *more* recent than Y”). The uncertainty around this issue is compounded by the fact that most previous studies have examined only one or two relational responses at a time (e.g., J. Hayes & Stewart, 2016; Vizcaíno-Torres et al., 2015).

Indeed, even measures assessing relational responding skills have tended to focus on only a small number of relational responses. The Relational Abilities Index (RAI, Colbert et al., 2017), for example, initially assessed only coordination, opposition, and quantity relations. However, researchers have now begun to recognise the theoretical and practical need for more elaborate assessments. Colbert and colleagues (2020), for instance, expanded the original RAI by adding distinction and temporal relations, as well as a more complex type of relational responding, analogical relations. Further, a recent study on the development of relational framing in children also assessed hierarchical responding in addition to those assessed by Colbert and colleagues (Kirsten & Stewart, 2021).

Identifying the similarities and differences between different relational responses represents a pressing issue. Current relational trainings may have inefficiencies whereby topographically different relations are trained separately despite them consisting of the same underlying functional class. If we could identify such underlying classes, then this information could be used to improve the efficiency of relational training. At the same time, at the theoretical level, it is of interest more generally to understand the precise ways in which different types of relational responses may be related to one another. Indeed, RFT explicitly makes such predictions regarding the interrelations of relational responses, but few (if any) empirical investigations have been undertaken (Hayes et al., 2001). Specifically, Hayes and colleagues (2001) stated that “one fairly clear prediction from RFT is that there should be some generalization of relational responding, particularly within families of relational frames. For example, an individual who learns to respond in accordance with sameness, may learn to respond in accordance with similarity (or opposition, since sameness is a combinatorially entailed aspect of opposition) more rapidly than, say, comparison. Similarly, learning across more closely associated families of relations may be more expected than learning across more distinct families” (p. 39).

Given the above, the current study aimed to investigate the potential generalisation of relational training effects among the different types of relational responding. To do this, we conducted two preregistered experiments which investigated the structure of relational responding. In a first experiment, we developed, and examined the correlations between different types of relational responding using an elaborated version of the RAI (Colbert et al., 2017, 2020). This elaborated RAI added three new relations to the assessment: containment relations (e.g., A is within B; B contains A), deictic relations (e.g., A is here now and B was there then; imagine A is B and B is A) and mathematical[1] relations (e.g., A + B is less than C + D). The correlations found in the first experiment then served as a guideline for our second experiment, in which we experimentally investigated whether training one type of relational responding would lead to greater generalised improvements in more related types compared to less related types of relational responding.

Experiment 1

In Experiment 1, we estimated the correlations among the different types of relational responses using an elaborated version of the RAI (see Materials section below for further information). Additionally, we also assessed the split-half and test-retest reliability of this elaborated RAI. To calculate the test-retest reliability of this measure, participants completed the RAI twice, with one week in-between. Only one study to date (Colbert et al., 2017) has examined the split-half and test-retest reliability of the RAI, and did so with a significantly smaller sample size (N = 35 compared to our N > 100).

[1] Notably, this is the first study to employ this relation, and “mathematical” relations are not typically discussed as canonical relations in RFT. As such, this subscale should be treated with relative caution.

Method

The data and materials for experiments 1 (https://osf.io/wmuyn/?view_only=84ddcc927c5a4147b95c8cf0d3f58207) and 2 (https://osf.io/sz39d/?view_only=df939676677147f79ba9be08b25b14cf) are available on the Open Science Framework. Both experiments were preregistered (Experiment 1: https://osf.io/xqdzp/?view_only=e837c7e11f4f4d488743997ddffd8a96; Experiment 2: https://osf.io/zp3xc/?view_only=9e34bdeb9b6d45d8b213c3bdbd3eddac)

Sample

Based on an a-priori power analysis, we aimed to collect complete data from 150 participants (see preregistration for detailed power analysis). Participants were tested via Prolific Academic (https://www.prolific.co/) and approved participants were paid at a rate of £7.50 per hour. In total, 218 participants were tested. After the exclusion of 73 participants (see Analysis section), the final sample consisted of 145 participants (105 female, 40 male, M_{age} = 28.27, SD_{age} = 6.09). Of these participants, 103 returned to complete the RAI one week later (73 female, 30 male, M_{age} = 28.50, SD_{age} = 6.41)[2].

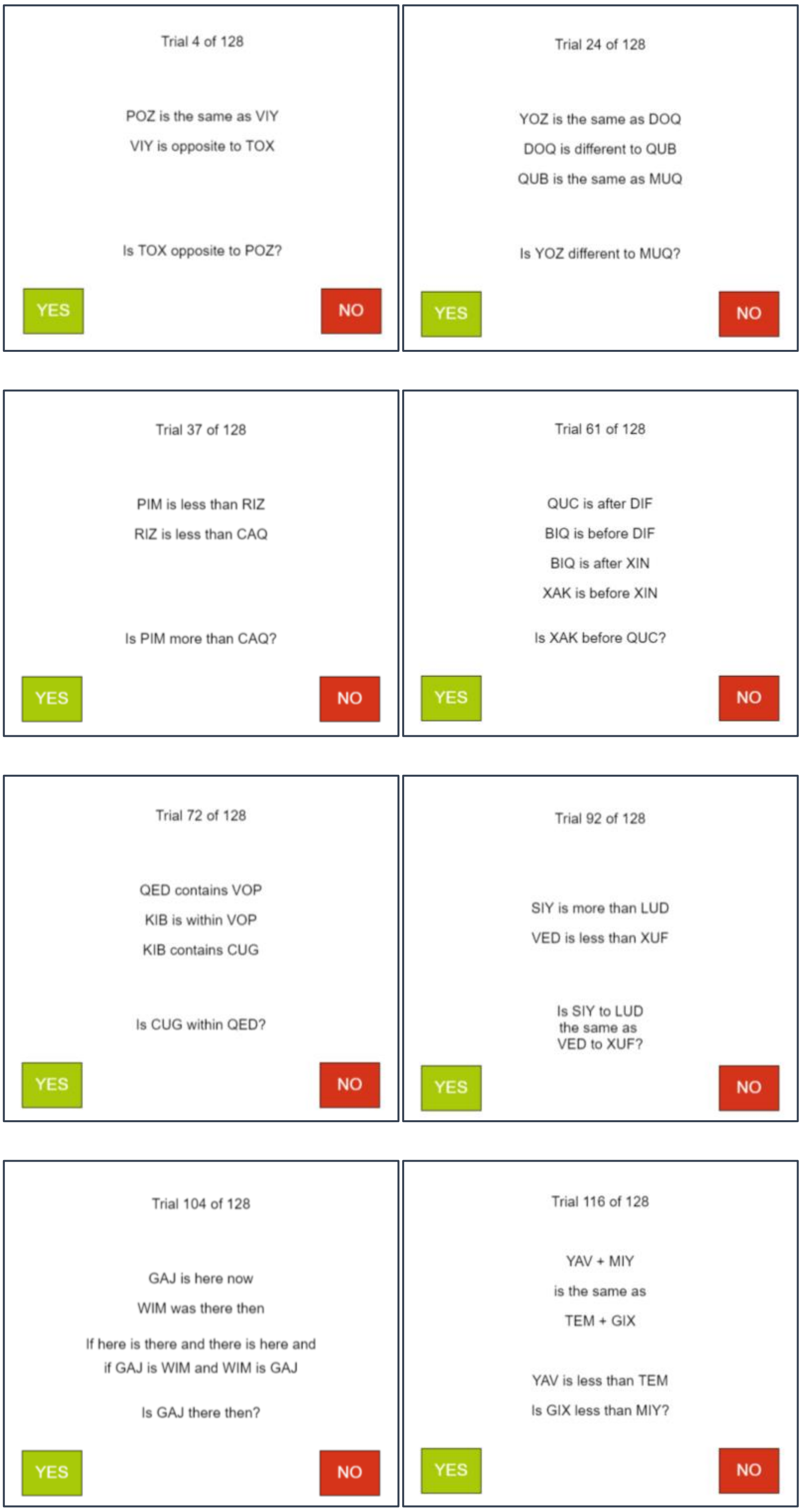
Materials

The Relational Abilities Index (RAI). The RAI used in this study contained 128 trials, divided across 8 different types of relations. The experiment was programmed in JavaScript, using the Lab.js online study builder (Henninger et al., 2019). Eight different relational subscales were used in this RAI; Figure 1 provides an example trial for each subscale. Trials of each relation type in the RAI increased progressively in difficulty along a series of dimensions, and participants were required to respond to each trial within a 30 second response window (for detailed information, see Colbert et al., 2020).

[2] We had expected a lower attrition rate from T1 to T2. However, with 103 participants this provides us with 95% power to detect a test-retest coefficient of r = .33. Given that the test-retest of RAI has previously been shown to be around .81 (Colbert et al., 2017), this sample size was more than sufficient.

Figure 1

Example trials for each subscale in the Relational Abilities Index



Note. The images follow the order of subscales in the RAI, starting from the upper left: opposition, difference, quantity, temporal, containment, analogy, deictic, and mathematical relations.

Procedure

In the first part of the study (timepoint 1; T1), participants first provided informed consent and demographic information (age and gender). After completing the RAI, which had a duration of approximately 30 minutes, they were asked to return to the Prolific site exactly one week later to complete the second part of the study. In the second part of the study (timepoint 2; T2), participants completed the RAI once again.

Analysis

Data Processing and Participant Exclusion. All data were processed and analysed using R (R Core Team, 2020) with the *tidyverse* package (Wickham et al., 2019). Seventy-three participants were excluded because of incomplete data on the RAI at T1 (39 participants) or for failing to meet our preregistered criteria; failing one or more of the attention checks (10 participants), having more than 20% of response times less than 5 seconds long (13 participants), or both (11 participants). This means that 145 participants were included in the analyses at T1. Forty-two additional participants were excluded from the analysis including T2 data, because they did not complete the RAI at T2 (28 participants) or because they showed signs of low effort responding, such as missing one or more attention checks (3 participants), having more than 20% of response times less than 5 seconds long (11 participants), or both (0 participants) during the RAI at T2. Therefore, the sample for the test-retest analysis contained 103 participants.

Due to a technical error, data for two trials in the quantity subscale were unusable. The score for quantity relations was thus based on 14 instead of 16 trials. After the removal of these trials, RAI scores at each timepoint were calculated based on the mean number of correct responses. Individual subscale scores were calculated similarly.

Results

Split-Half Reliability of the Overall and Subscale RAI Scores at T1.

We computed the split-half reliability for the RAI using an odd-even approach on the T1 data (*N* = 145). The Spearman-Brown corrected split-half reliability of the full RAI was high, $r(143) = .90$, 95% CI [.86, .93]. Table 1 provides an overview of the split-half reliabilities for all eight subscales of the RAI.

Test-Retest Reliability of the Overall and Subscale RAI Scores From T1 to T2.

We computed the test-retest reliability by correlating the RAI scores at T1 with the scores at T2, both for overall and subscale RAI scores (*N* = 103). The test-retest reliability of the full RAI was high, $r(101) = .85$, 95% CI [.79, .90]. However, the performance of the subscales was much more varied in this regard. Table 2 provides an overview of the test-retest reliabilities for each of the eight subscales.

Correlations Between RAI Trial Types at T1.

This analysis was performed on the T1 data (*N* = 145). We computed Pearson correlations between each of the 8 trial types. We computed two confidence intervals for each correlation: the first was based on the raw estimates, while the second was adjusted with the Holm method. The decision to determine whether two correlations differed was based on the adjusted confidence intervals. If the adjusted CI for one scale did not contain the estimate of another, we decided that the correlations differed significantly. We found that all but two of the correlations among the subscales were positive and significant. Two correlations, namely between the analogy and opposition subscale, $r(143) = .10$, $p = .233$, 95% CI_{adj} [-.16, .35], and between the analogy and difference subscale, $r(143) = .11$, $p = .192$, 95% CI_{adj} [-.15, .36], were not statistically significant. Table 3 provides an overview of all correlations and their raw and Bonferroni-adjusted 95% confidence intervals.

Table 1. Split-half reliability of the subscales of the RAI at T1.

Subscale	Split-Half Reliability	95% CI
1. Opposition	.52	[.34, .66]
2. Difference	.08	[-.27, .34]
3. Quantity	.72	[.61, .80]
4. Temporal	.77	[.68, .83]
5. Containment	.80	[.72, .85]
6. Analogy	-.61	[-1.00, -.16]
7. Deictic	.73	[.62, .80]
8. Mathematical	.75	[.65, .82]

Table 2. Test-retest reliability of the subscales of the RAI from T1 to T2.

Subscale	Test-Retest Reliability	95% CI
1. Opposition	.31	[.13, .48]
2. Difference	.68	[.56, .77]
3. Quantity	.64	[.50, .74]
4. Temporal	.71	[.60, .80]
5. Containment	.73	[.63, .81]
6. Analogy	.38	[.20, .53]
7. Deictic	.59	[.45, .70]
8. Mathematical	.59	[.45, .71]

Table 3. The correlations with raw and adjusted confidence intervals among the different subscales of the Relational Abilities Index at T1.

Subscale	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1. Opposition	-						
2. Difference	.37** [.22, .50] CI_{adj} = [.13, .57]	-					
3. Quantity	.37** [.22, .50] CI_{adj} = [.12, 57]	.45** [.31,.57] CI_{adj} = [.22, .63]	-				
4. Temporal	.36** [.21, .49] CI_{adj} = [.11, .56]	.47** [.34, .59] CI_{adj} =[.25, .65]	.56** [.44, .66] CI_{adj} = [.36, .72]	-			
5. Containment	.27** [.11, .41] CI_{adj} = [.01, .49]	.33** [.18, .47] CI_{adj} = [.08, .54]	.39** [.24, .52] CI_{adj} = [.15, .59]	.59** [.47, .68] CI_{adj} = [.39,.73]	-		
6. Analogy	.10 [-.06, .26] CI_{adj} = [-.16, 35]	.11 [-.05, .27] CI_{adj} = [-.15, .36]	.37** [.22, .50] CI_{adj} = [.12, .57]	.31** [.16, .45] CI_{adj} = [.06, .53]	.19* [.03, .34] CI_{adj} = [-.07, .43]	-	
7. Deictic	.39**[.24, .52] CI_{adj} = [.14, .58]	.25** [.09, .39] CI_{adj} = [-.01, .47]	.37** [.22, .51] CI_{adj} = [.13, 57]	.32** [.16, .46] CI_{adj} = [.07, .53]	.31** [.15, .45] CI_{adj} = [.06, .52]	.28** [.12, .43] CI_{adj} = [.03, .50]	-
8. Mathematical	.23**[.07, .38] CI_{adj} = [-.02, .46]	.17* [.00, .32] CI_{adj} = [-.09, .41]	.26** [.11, .41] CI_{adj} = [.01, .49]	.42** [.28, .55] CI_{adj} = [.18, .61]	.37** [.22, .50] CI_{adj} = [.12, .57]	.31** [.16, .45] CI_{adj} = [.06, .53]	.39** [.24, .52] CI_{adj} = [.15, .59]

Note. Values within the square brackets indicate the 95% confidence interval for each correlation. CI_{adj} indicates the 95% confidence interval adjusted with the Holm method. * indicates that p<.05 and ** indicates p<.01. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg method.

Discussion

In Experiment 1, we calculated the correlations among the different subscales and assessed the reliability of our elaborated RAI as a measure of relational responding. We found that our RAI was a reliable measure of relational responding, showing high test-retest and split-half reliability, but with substantial variation across the different subscales. Second, we found that almost all types of relational responding were positively and significantly correlated to varying degrees. Only two correlations, namely the correlation between analogical and opposition relations and analogical and difference relations, were not statistically significant. With this information in mind, we moved on to our second experiment, wherein we examined the generalisation of training one type of relation to both strongly- and weakly-related relations.

Experiment 2

The aim of Experiment 2 was to experimentally examine the interrelatedness of different types of relational responses: namely, whether training relational responding in accordance with one relation frame generalised to related relations, and whether generalisation proportionate to the relatedness of relations could be observed. We therefore trained two relations in two separate groups of participants[3]. Specifically, one group was trained on difference relations, while the other group was trained on containment relations. We chose to train these relations for two reasons. First, they allowed us to construct comparable type-specific relational training, which was much more difficult to do for

[3] After data collection for Experiment 2 had started, we discovered an error in the scoring of three trials in the RAI used in Experiment 1. Further, a technical error with two trials of the quantity subscale (mentioned in Experiment 1) was also discovered. This affected most of our estimated correlations in Experiment 1, which in turn affected our decisions for Experiment 2. Because data collection for Experiment 2 had already started, we could not change the to-be-trained relations. However, we could (and did) change which relations we analysed accompanying the trained relations. Therefore, the specific relations we analysed differed from what we had originally preregistered for Experiment 2 (https://osf.io/zp3xc/?view_only=9e34bdeb9b6d45d8b213c3bdbd3eddac, see the deviations from preregistration document). For our analyses, we corrected the scoring of the three trials which were scored incorrectly and removed the two trials in the quantity subscale which were unusable.

compound relations, such as mathematical/analogical/deictic relations. Second, the initial correlation between these relations in Experiment 1 (after excluding compound relations) was descriptively the weakest. Although after reanalysing the data of Experiment 1 (see footnote 1) this correlation was no longer descriptively the weakest, it did not significantly differ from the weakest correlation (i.e., between containment and opposition relations).

We had three predictions regarding our second experiment. First, we predicted that performance on the trained relation would increase after the relation-specific training (RQ1). Second, we predicted that the relation-specific training would also increase performance on types of relational responding that were strongly correlated with the trained relation based on Experiment 1's correlations (RQ2). Both containment and difference relations were most strongly correlated with temporal relations. Thirdly and most importantly, we predicted that we would observe generalisation of training effects proportionate to the relatedness of other relations, such that the generalisation of training would be greater for types of relational responding that were strongly, compared to weakly, related to the trained relation (RQ3). Based on the results of our first experiment, we chose to compare the effect of both containment and difference training on temporal (strongly related) vs. analogical (weakly related) relations. We predicted a greater increase in the performance on temporal compared to analogical relations after both containment and difference training.

Method

Sample

An a-priori power analysis (which utilised coefficients extracted from Experiment 1) indicated that a sample size of 85 participants per condition (i.e., 170 participants in total) would provide us with 98% power to detect a small-to-medium Cohen's d effect size in RQ1 and 91% power to detect a medium interaction effect for RQ2. Furthermore, this would lead to 82% and 83% power for the within-subjects ANOVA of the difference and containment trainings, respectively, for RQ3. Therefore, we endeavoured to collect completed data from 170 participants.

Data was collected via Prolific. All participants were between the ages of 18 to 40, had a 95% approval rating for previous studies on Prolific, and no participation in previous studies from our research group. In total, 335 participants started the first assessment of the experiment. After the rejection of 158 participants (see Analyses section for details), our final sample for the analyses for RQ1 and RQ2 consisted of 177 participants (134 female, M_{age} = 30.14, SD_{age} = 6.51), with 89 participants in the containment training condition (66 female, M_{age} = 29.36, SD_{age} = 5.81) and 88 participants in the difference training condition (68 female, M_{age} = 30.93.04, SD_{age} = 7.09). For the confirmatory analyses of RQ3, however, one additional participant was excluded, due to incomplete data for the analogy subscale in the post-training assessment of the RAI (N = 176). For a subsequent exploratory analysis of RQ3, two extra participants were excluded due to incomplete data for the mathematical subscale in the post-training assessment of the RAI (N = 174).

Design

We investigated the effects of difference and containment training using a mixed between-within design, with trained relation (difference relations, containment relations) as the between-subject factor and the timepoint (T1, T2) as the within-subject factor. Our dependent variable of interest was the difference in the accuracy of the responses on the RAI subscales at T1 (before training) versus T2 (after training).

Materials

The Relational Abilities Index (RAI). The RAI was identical to Experiment 1.

Relational Training. The difference and containment trainings were created bespoke for the current experiment and were based on the SMART programme (Cassidy et al., 2011; see Supplementary Materials for further information). Each training consisted of 16 stages, which increased progressively in difficulty along a series of dimensions: the number of premises (2, 3, or 4), the relations in the premises (same or mixed for difference training; contains, is within, or mixed for containment training), the direction of the premises (forwards or mixed; i.e., need for mutual entailment), the relation in the question (same, different, or mixed for difference training; contains, is within, or mixed for containment training), and the nodal distance (0, 1, 2, or 3; i.e., the complexity of combinatorial entailment; see Cassidy et al., 2016 for more information about these dimensions). Tables 1 and 2 in the Supplementary Materials provide an overview of all 16 stages in the difference and containment training respectively.

Procedure

Because this study was labour-intensive, we sampled in batches. We first conducted a pilot run and then continued with four runs of participants. Each batch took approximately one week.

At the beginning of each run, participants signed the informed consent and provided demographic information (age and gender). Then, they completed the RAI. During the week which followed, participants completed three training sessions of 30 minutes each. Half of the participants were trained on difference relations, while the other half was trained on containment relations. Assignment to trainings was quasi-random, with both groups matched on overall baseline relational abilities. One day after their last training session, participants completed the RAI again.

Analysis

Data Processing and Participant Exclusion

All data were processed and analysed using R (R Core Team, 2020) with the *tidyverse* package (Wickham et al., 2019). We initially tested 335 participants at T1. One hundred and nine participants were excluded because of incomplete data on the RAI at T1 (76 participants) or for failing to meet our preregistered criteria; failing one or more of the attention checks (1 participant), having more than 20% of response times less than 5 seconds long (22 participants), or both (10 participants). This meant that 226 participants were included in the training. Forty-three participants were excluded because they did not complete all training sessions (40 participants) or because they did not pass the first stage of training (3 participants). Five participants were additionally excluded at T2, because of incomplete data (1 participant) or because of missed attention checks (1 participant), having more than 20% of response times less than 5 seconds long (2 participants), or both (1 participant).

Results

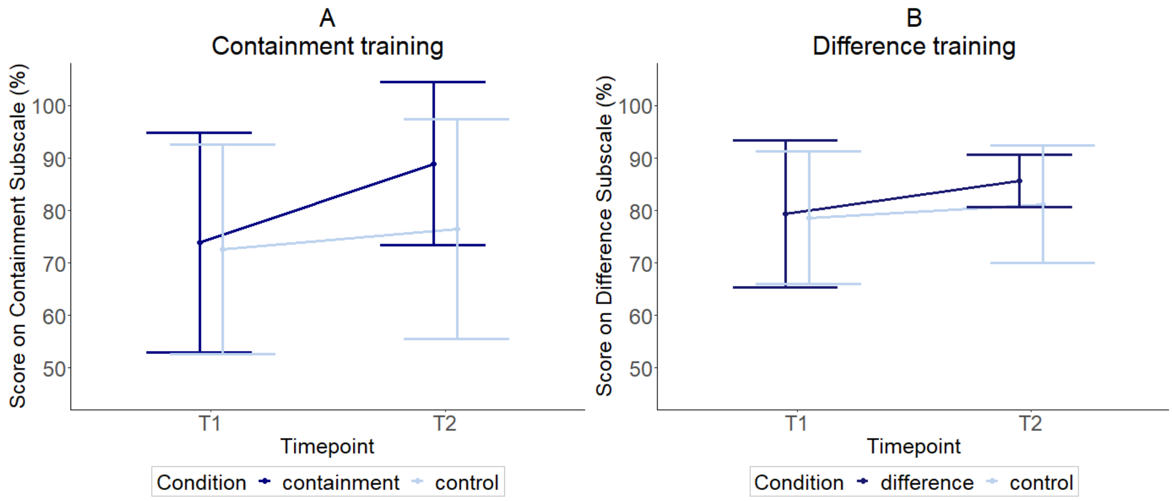
In line with our preregistration, if our Frequentist analyses did not reveal a significant effect, we computed a Bayesian multilevel equivalent of the corresponding model (using uninformed priors and modelling participant as a random effect) to quantify evidence for the absence of an effect (Schmalz et al., 2021).

Confirmatory Analyses

H1. Performance Increase on the Trained Relation After Training. To test whether there was an increase in performance on the trained type of relational responding, we conducted a one-sided paired t-test on the RAI scores for the trained relation within each condition before (T1) and after (T2) training. For containment training (Figure 2A), participants scored significantly higher on the containment subscale after (M = 88.90%, SD = 15.53%) than before containment training (M = 73.90%, SD = 20.93%), $t(88) = 7.39$, $p < .001$, $d_z = 0.78$. Likewise, for difference training (Figure 2B), participants scored significantly higher on the difference subscale after (M = 85.65%, SD = 5.03%) compared to before (M = 79.39%, SD = 14.00%) difference training, $t(87) = 4.63$, $p < .001$, $d_z = 0.49$.

Figure 2

Performance change on trained relation for Containment (A) and Difference (B) training



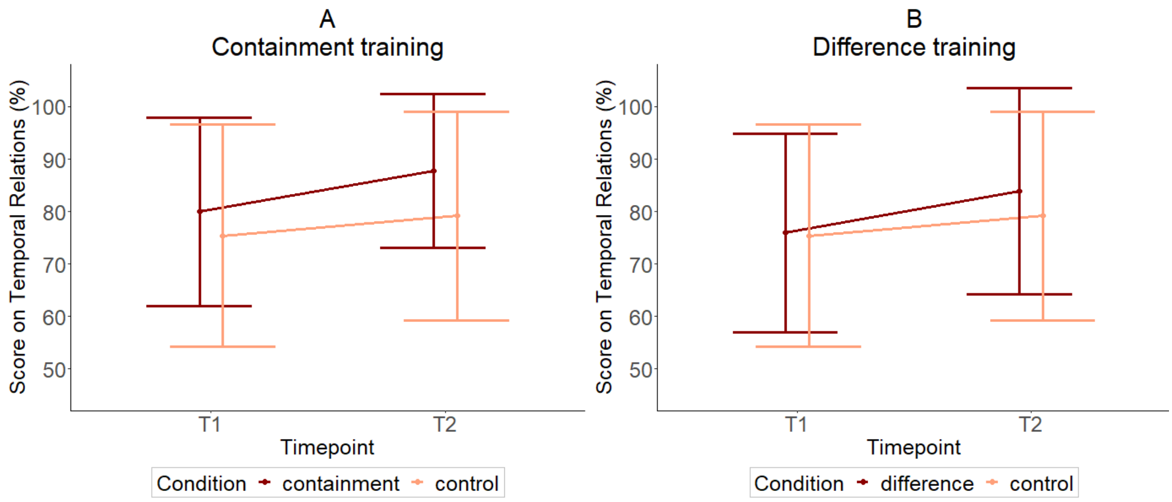
Note. The error bars represent the standard deviation of the mean accuracy. The light blue lines represent the accuracy for the same subscale in Experiment 1 (i.e., related to exploratory RQ1).

We next investigated whether the magnitude of these training effects differed between experimental conditions. To do this, we conducted a mixed between-within ANOVA, with the trained relation (difference, containment) as a between-subject factor and timepoint (T1, T2) as a within-subject factor. In line with the above t-tests, we found a significant main effect of timepoint, $F(1,175) = 75.66$, $p < .001$, $\eta_p^2 = .30$, with higher scores at T2 (M = 87.29%, SD = 11.65%) than T1 (M = 76.63%, SD = 17.98%). However, we also found a significant interaction between condition and timepoint, $F(1,175) = 12.81$, $p < .001$, $\eta_p^2 = .07$, with a larger training effect for containment training ($M_{increase}$ = 15.01%, $SD_{increase}$ = 19.17 %) compared to difference training ($M_{increase}$ = 6.26%, $SD_{increase}$ = 12.67 %).

H2. Generalisation of Training Effects to a Strongly Correlated Relation. In each training group, we conducted a one-sided paired t-test on the RAI scores before (T1) and after training (T2) for the relation type that correlated most strongly with the trained relation (i.e., temporal relations in both cases). We found the expected generalisation for both containment training, $t(88) = 4.02$, $p < .001$, $d_z = 0.43$, and for difference training, $t(87) = 4.11$, $p < .001$, $d_z = 0.44$ (see Figure 3).

Figure 3

Performance change on temporal relations for Containment (A) and Difference (B) training

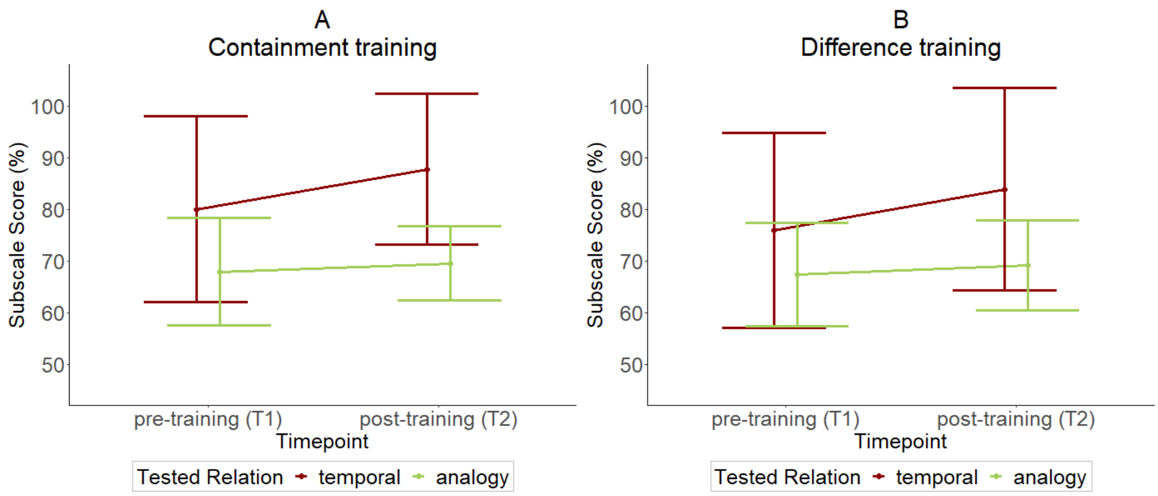


Note. The light red lines represent the accuracy change for the temporal subscale in Experiment 1 (i.e., related to exploratory RQ2).

H3. Greater Generalisation of Training Effects for Strongly Compared to Weakly Correlated Relations. To investigate the third research question, we compared the performance before (T1) and after training (T2) for the relations that showed the strongest versus weakest correlation with the trained relation. For both containment and difference training, this consisted of temporal relations as the most strongly related, and analogical relations as the most weakly related. Most critically, we observed the expected interaction between timepoint and relation type, such that greater improvements in temporal relations compared to analogical relations were seen for both containment training, $F(1,88) = 8.09$, $p = .006$, $\eta_p^2 = .08$, and for difference training, $F(1,86) = 6.89$, $p = .010$, $\eta_p^2 = .07$ (see Figure 4).

Figure 4

Increase on temporal compared to analogical relations for Containment (A) and Difference (B) training



Exploratory Analyses

In addition to our preregistered confirmatory analyses, we also conducted a series of non-preregistered exploratory analyses to unpack our research questions further.

RQ1. Performance Increase on the Trained Relation After Training Compared to Control. To exclude the possibility that the increase in performance on the trained relation was caused by test-retest effects, we compared the performance difference for the trained subscale before versus after training to the performance difference for this subscale between T1 and T2 in Experiment 1. In doing so, we could determine whether observed changes in performances on the measures were a result of our training intervention or merely due to natural test-retest effects (i.e., since Experiment 1 examined test-retest effects without any additional intervention). To do this, we conducted two between-within ANOVA with timepoint (T1, T2) and condition (control vs. training) as factors: one for containment training, and one for difference training. We found significant interaction effects for both containment training, $F(1,190) = 20.37$, $p < .001$, $\eta_p^2 = .10$, and difference training, $F(1,189)$

= 5.29, p = .022, η_p^2 = .03, indicating that the impact of training observed in Experiment 2 was not merely due to test-retest effects (see Figure 2).

RQ2. Generalisation of Training Effects to a Strongly Correlated Relation Compared to a Control. We also added a post-hoc, exploratory analysis to test whether the performance increase on the closely related relation (i.e., temporal relations) after both containment and difference training differed from the test-retest of Experiment 1. Because the relation to be compared was the same for both training conditions (i.e., temporal relations), we combined the data for both conditions into one “experimental condition” to acquire a larger sample size (N = 177), which we then compared to the data from Experiment 1 (N = 103) to rule out test-retest effects. We conducted a between-within ANOVA with timepoint (T1, T2) and condition (control vs. experimental) as within- and between-subject factors respectively. Importantly and against our expectations, the condition x timepoint interaction was not significant, $F(1, 278) = 3.72$, p = .055, η_p^2 = .01, which suggests that the increase in the scores on the temporal subscale in the experimental conditions (M = 7.87%, SD = 18.19%) did not differ significantly from the increase in the control condition (M = 3.74%, SD = 15.69%). However, the Bayesian analysis suggested that the presence of an interaction was very substantially more likely than its absence (BF_{10} = 8515).

RQ3. Greater Generalisation of Training Effects to Strongly Compared to Weakly Correlated Relations. We conducted an exploratory between-within ANOVA to test our third research question. From our Experiment 1 data, we found that opposition relations were descriptively (though not significantly) more related to difference relations than containment relations. Likewise, we observed that mathematical relations were descriptively (though not significantly) more related to containment relations compared to difference relations. If relational training generalises proportionate to the relatedness of relations, then we would expect containment training to increase mathematical relations to a greater extent than opposition relations, and we would expect the opposite pattern for difference training. Therefore, we conducted a three-way mixed ANOVA, with RAI subscale score as the DV, timepoint (T1 vs. T2) as one IV, relational subscale (opposition vs. mathematical) as another IV, and relational training (containment vs. difference) as the third IV. We would expect to find a three-way interaction effect, whereby improvements in opposition relations from T1 to T2 are greater for difference training compared to containment training, whereas improvements in mathematical relations are greater for containment training compared to difference training. Although we found a significant interaction between condition and timepoint, $F(1,172) = 4.42$, $p = .037$, $\eta_p^2 = .03$ (indicating that the relational training in general improved relational abilities), the three-way interaction effect was not significant, $F(1,172) = 0.84$, $p = .359$, $\eta_p^2 < .01$. The Bayesian analysis ($BF_{10} = 0.79$) indicated neither support for nor against the presence of an interaction effect.

Discussion

In Experiment 2, we investigated whether training one type of relational responding also increased performance on related types of relational responding. In general, we found evidence which supported the idea that relational training can improve relational abilities, both for directly trained relations as well as non-trained relations. We also found evidence that improvements in non-trained relations vary as a function of how related the non-trained relation is to the directly trained relation.

General Discussion

Overview of the Findings

In Experiment 1, we found that different types of relational responding were indeed related, as almost all subscales in our elaborated RAI correlated positively across participants. Importantly, we found that our elaborated RAI was an overall reliable measure of relational responding, with high split-half and test-retest reliability (though notably the individual subscales performed worse in this regard). Experiment 2 built on these findings within an experimental paradigm. More precisely, we tested whether generalisation of training effects would be greater for closely related than distantly related relations. We found that specific types of relational responding can indeed be trained directly using our specially designed 3-session training. Our short-term training increased performance on a highly correlated subscale, temporal relations. Moreover, in line with our predictions, the performance increase after training was greater for this subscale than for the analogy subscale, which was only weakly correlated with the trained relations. Our post-hoc exploratory analyses did not show a significantly greater effect on the performance for temporal relations after containment or difference training compared to the test-retest effect in Experiment 1. However, our Bayesian multilevel analysis suggested that the presence of an interaction effect was substantially more likely than its absence. For the exploratory analysis of RQ3, we found no significant effect in the Frequentist analysis and inconclusive evidence from our Bayesian analysis. Given that we did not power for these exploratory analyses, no firm conclusions can be drawn from them; particularly from the latter analysis of RQ3, given that three-way interactions typically require rather large sample sizes for sufficient power.

Theoretical and Practical Implications

Our results have several critical theoretical and practical implications for RFT and our understanding of relational responding. Firstly, on a practical level, our results indicated that relational abilities can be improved (at least in the short term) using relatively brief, 3-session relational training interventions. Indeed, although previous results training stimulus equivalence responding have indicated similar improvements after 3-session training (Cummins et al., 2018), typically relational training interventions are much more elaborate and time-consuming. Our results and newly developed mini-training therefore illustrate a path to conduct basic empirical research on relational training in a much more efficient manner. As well as this, the RAI used in this study represents one of the most elaborate computerised assessment of relational responding created to date, further expanding the original RAI (Colbert et al., 2017, 2020). This RAI is also available as open-source software (https://github.com/JamieCummins/relational-abilities-index).

At the theoretical level, our results also provide some critical additions to the literature. Firstly, our studies provide, for the first time, an elaborate overview of the correlations between different types of relational responding. Our second experiment also affirms a core assumption of RFT: namely, that different types of relational responses are differentially related to one another (S. C. Hayes et al., 2001). Our study is the first, further, to demonstrate that manipulating one relational response differentially impacts performances on related relations as a function of the correlation between these relations. This represents a first step towards precisely unpacking the interrelations between different types of relational responding. Of course, it remains unclear presently how best to characterise these interrelations. It is not clear, for example, whether it is more accurate and useful to consider relational responding as a single operant skill (as predicted by theories such as Process Overlap Theory; Kovacs & Conway, 2016), or as a collection of several distinct operants. To answer this more thoroughly, a larger sample and alternative statistical methods (e.g., factor analysis, principal components analysis, or network analysis) are required. However, if similar studies such as this are conducted in the future, then our data could be pooled with these studies to directly investigate this with appropriate power.

The use of methods such as factor analysis to identify clusters of relational operant responses would represent a data-driven perspective on understanding the interrelations between relational responses. However, it is important to note that existing theoretical frameworks within RFT may also shed light on this. Indeed, as one reviewer noted, the

HDML framework may be used as a source of inspiration for the continuation of this line of work (Harte et al., 2020). The approach used here may interface well with the HDML's levels of analysis of arbitrarily applicable relational responding (AARRing; e.g., coherence, derivation) which may in turn be used as a source of inspiration for future such studies. This could provide insight into the contexts under which the observed effects do and do not occur, which would provide greater contextualistic information regarding the relational responding observed here. We return to this point later in the below section.

Limitations

Both experiments in the current study had some limitations. For instance, our elaborated RAI assessed participants' relational responding using only 16 trials per subscale (14 trials for quantity relations), leading to poor estimation around subscale scores. Indeed, the adjusted confidence intervals at the group level often included a range of more than .4. This also led to high variability between subscales in terms of their reliability. Therefore, only large differences between estimated correlations reached significance. This may well have affected our ability to detect differential generalisation effects, particularly within analyses that were otherwise poorly powered (i.e., the exploratory analyses).

The analogy subscale was the only subscale with a negative value for the split-half reliability, in addition to its low test-retest reliability. A possible explanation for the low reliability of this subscale is that it contained too much variation across trials. Namely, not only the difficulty but also the relations within the premises differed across the trials of this subscale. The low reliability of both the opposition and analogy subscale might also explain why these subscales showed no significant correlation, although the analogy subscale included opposition relations in the premises and the questions. We did not consider the reliability of individual subscales when we selected relations for our second experiment, which is why all three subscales mentioned above were also included, either as the trained relation or outcome measures. However, reliability, especially test-retest reliability, of these subscales is important to measure training effects. For these reasons, improving the reliability of subscales in the relational assessment is critical to future work seeking to make inferences regarding the structure of relational responding. If we cannot reliably measure difference relational subscales, then making inferences regarding these subscales with sufficient statistical power will be extremely difficult.

In our second experiment, the comparison between temporal and analogical relations was not optimal. The analogy subscale not only had poor reliability, but our results also indicated that participants had overall lower performance on this subscale compared to the temporal subscale. This suggests that the analogy subscale was more difficult than the temporal subscale, which could have caused the analogy subscale to be somewhat resistant to improvement, unrelated to its correlation with the trained relations. This could then potentially be an alternative explanation for the difference in generalisation found in Experiment 2. However, this seems unlikely given that our exploratory analysis for RQ2 indicated that a timepoint x condition interaction was substantially more likely than its absence. Regardless, it is important to consider potential differences in both the difficulty as well as the reliability of included relations for subsequent studies.

Beyond issues of psychometrics, it is also important to bear in mind that the act of naming a subscale (e.g., as “mathematical”, “deictic”, etc.) is an imposition by the experimenter, and subscales may not be as discrete or clear as they may seem in this regard. For example, as one reviewer flagged, the “deictic” subscale of the RAI bears similarity to the perspective-taking protocol of McHugh et al. (2004), which was developed specifically to assess perspective-taking in children. However, the validity of this protocol in adults has recently been questioned (Kavanagh et al., 2020). More generally, it is important to bear in mind that there is need for external validation of the meaningfulness of these individual subscales with related measures, particularly those which are more “high level” such as deictic responding.

Future Research

Future studies should aim to refine the correlational patterns found in Experiment 1 after improving upon the psychometric properties of the RAI subscales. Detecting subtle differences in correlations would not only have theoretical value but would also allow us to improve the method for studies such as our second experiment. That is, a more fine-grained pattern would allow researchers to conduct the exploratory analysis of RQ3 based on relations that show significantly different correlations with the trained relations.

The results of our second experiment represent an important step towards understanding the dynamics of learning at play in relational training. However, this understanding can and should be forwarded to a greater extent by future studies. On the one hand, it is important to note that most relational training interventions to date tend to focus almost exclusively on control over the relations between stimuli and less on the control over the specific functions related between those stimuli (Delabie et al., 2022). Incorporating such control into future relational trainings could allow CBS researchers to then investigate whether different relational responses can be experimentally controlled to be differentially related to one another in a similar experimental approach as to the current one.

One reviewer noted that it may also be of interest to examine the generalisation of effects observed in the current paradigm to other aspects of relational responding. Here we examined the accuracy of participants responding to relational syllogisms, but it may also be of interest to examine whether response features such as flexibility can also be affected by this training approach. For example, if Relation 1 is trained, does *flexibility* on Relation 1 improve (O’Toole et al., 2009)? Further still, if Relation 2 is strongly correlated to Relation 1, and Relation 3 is weakly correlated, will flexibility in Relation 2 responding improve to a greater extent than flexibility in Relation 3? Further still, it may be of interest to incorporate elements from the HDML in this regard (Harte et al., 2020). For instance, if the levels of derivation (i.e., the extent to which a particular pattern of derived relational responding has previously been "practiced" or emitted) in the training and testing procedures were manipulated, might this have an impact on the observed correlations between difference relational responses? Similar questions may also be asked of other features of the relational responses, such as coherence and complexity (see Harte et al., 2020, for a discussion of these concepts). Indeed, if differences in the relationship between different relational responses are observed as a function of manipulating any of the above variables, then this would represent a highly useful approach to examining the dynamics of AARRing while also accounting for the contextualistic features which impact these responses.

One further approach which may be of interest would be to replicate the approach taken in this paper (namely, to use the correlational structure of a first experiment to influence the experimental design of a second experiment) while also including other measures, such as measures of cognitive ability. For instance, if we would find that a measure of fluid intelligence, such as Raven's Matrices, were strongly correlated with one type of relational response (e.g., difference responding), and that a measure of perspective-taking were strongly correlated with another type of relational response (e.g., deictic responding), then it may be of interest to examine whether a similar experimental design to Experiment 2 can also produce expected effects on these other measures (i.e., examining if training difference responding improves fluid intelligence more than training deictic responding, and whether the opposite pattern of results is observed for perspective-taking). Such a study would provide even further insight into the dynamics of relational responding, as well as a more fine-grained view on the impacts of relational training on other outcomes.

Conclusion

To conclude, we found interrelations among the different types of relational responding, in line with the predictions of RFT, and initial evidence for a generalisation of training effects to other untrained relations which is proportion to the relatedness of those relations (although this evidence is couched in uncertainty due to the variability in measurement properties of the RAI subscales). These results provide novel insights into the dynamics of relational responding, as well as illustrating a path towards a more functional understanding of different relational operant responses.

References

Amd, M., & Roche, B. (2018). Assessing the effects of a relational training intervention on fluid intelligence among a sample of socially disadvantaged children in Bangladesh. *Psychological Record*, *68*(2), 141–149. https://doi.org/10.1007/s40732-018-0273-4

Barnes-Holmes, Y., Barnes-Holmes, D., & Smeets, P. M. (2004). Establishing relational responding in accordance with opposite as generalized operant behavior in young children. *International Journal of Psychology and Psychological Therapy*, *4*(3), 559–586.

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

Barnes-Holmes, Y., McHugh, L., & Barnes-Holmes, D. (2004). Perspective-taking and

Theory of Mind: A relational frame account. *The Behavior Analyst Today*, *5*(1), 15–25. https://doi.org/10.1037/h0100133

Cassidy, S., Roche, B., Colbert, D., Stewart, I., & Grey, I. M. (2016). A relational frame skills training intervention to increase general intelligence and scholastic aptitude. *Learning and Individual Differences*, *47*, 222–235. https://doi.org/10.1016/j.lindif.2016.03.001

Cassidy, S., Roche, B., & Hayes, S. C. (2011). A relational frame training intervention to raise intelligence quotients: A pilot study. *Psychological Record*, *61*(2), 173–198. https://doi.org/10.1007/BF03395755

Cassidy, S., Roche, B., & O'Hora, D. (2010). Relational Frame Theory and human intelligence. *European Journal of Behavior Analysis*, *11*(1), 37–51. https://doi.org/10.1080/15021149.2010.11434333

Colbert, D., Dobutowitsch, M., Roche, B., & Brophy, C. (2017). The proxy-measurement of intelligence quotients using a relational skills abilities index. *Learning and Individual Differences*, *57*, 114–122. https://doi.org/10.1016/j.lindif.2017.03.010

Colbert, D., Malone, A., Barrett, S., & Roche, B. (2020). The Relational Abilities Index+: Initial validation of a functionally understood proxy measure for intelligence. *Perspectives on Behavior Science*, *43*(1), 189–213. https://doi.org/10.1007/s40614-019-00197-z

Colbert, D., Tyndall, I., Roche, B., & Cassidy, S. (2018). Can SMART training really increase intelligence? A replication study. *Journal of Behavioral Education*, *27*(4), 509–531. https://doi.org/10.1007/s10864-018-9302-2

Cummins, J., Roche, B., Tyndall, I., & Cartwright, A. (2018). The relationship between differential stimulus relatedness and implicit measure effect sizes. *Journal of the Experimental Analysis of Behavior*, *110*(1), 24-38. https://doi.org/10.1002/jeab.437

Delabie, M., Cummins, J., Finn, M., & De Houwer, J. (2022). Differential Crel and Cfunc acquisition through stimulus pairing. *Journal of Contextual Behavior Science*, *24*, 112-119. https://doi.org/10.1016/j.jcbs.2022.03.012

Dixon, M. R. (2014). *PEAK relational training system: Direct training module*. Shawnee Scientific Press.

Dougher, M. J., Hamilton, D. A., Fink, B. C., & Harrington, J. (2007). Transformation of the discriminative and eliciting functions of generalized relational stimuli. *Journal of the Experimental Analysis of Behavior*, *88*(2), 179–197. https://doi.org/10.1901/jeab.2007.45-05

Dunne, S., Foody, M., Barnes-Holmes, Y., Barnes-Holmes, D., & Murphy, C. (2014). Facilitating repertoires of coordination, opposition distinction, and comparison in young children with autism. *Behavioral Development Bulletin*, *19*(2), 37–47. https://doi.org/10.1037/h0100576

Dymond, S., & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *Behavior Analyst*, *23*(2), 239–254. https://doi.org/10.1007/BF03392013

Gore, N. J., Barnes-Holmes, Y., & Murpy, G. (2010). The relationship between intellectual functioning and relational perspective-taking. *International Journal of Psychology & Psychological Therapy*, *10*(1), 1–17. https://psycnet.apa.org/record/2010-05335-001

Harte, C., Barnes-Holmes, D., Barnes-Holmes, Y., & Kissi, A. (2020). The study of rule-governed behavior and derived stimulus relations: Bridging the gap. *Perspectives on Behavior Science*, *43*(2), 361-385.

Hayes, J., & Stewart, I. (2016). Comparing the effects of derived relational training and computer coding on intellectual potential in school-age children. *The British Journal of Educational Psychology*, *86*(3), 397–411. https://doi.org/10.1111/bjep.12114

Hayes, J., Stewart, I., & McElwee, J. (2016). Assessing and training young children in same and different relations using the Relational Evaluation Procedure (REP). *Psychological Record*, *66*(4), 547–561. https://doi.org/10.1007/s40732-016-0191-2

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-skinnerian account of human language and cognition*. Kluwer Academic/Plenum Publishers.

Hayes, S. C., & Hayes, L. J. (1992). Verbal relations and the evolution of behavior analysis. *American Psychologist*, *47*(11), 1383–1395. https://doi.org/10.1037/0003-066X.47.11.1383

Healy, O., Barnes-Holmes, D., & Smeets, P. M. (2000). Derived relational responding as generalized operant behavior. *Journal of the Experimental Analysis of Behavior*, *74*(2), 207–227. https://doi.org/10.1901/jeab.2000.74-207

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P., & Hilbig, B. (2019). *lab.js: A free, open, online study builder*. https://doi.org/10.31234/osf.io/fqr49

Heyes, C. (2018). *Cognitive Gadgets: The cultural evolution of thinking*. Harvard University Press. https://doi.org/10.4159/9780674985155

Hughes, S., & Barnes-Holmes, D. (2016). Relational Frame Theory: The basic account. In *The Wiley Handbook of Contextual Behavioral Science* (pp. 129–178). https://doi.org/10.1002/9781118489857.ch9

Kirsten, E. B., & Stewart, I. (2021). Assessing the development of relational framing in young children. *The Psychological Record*, 1–26. https://doi.org/10.1007/s40732-021-00457-y

Lipkens, R., Hayes, S. C., & Hayes, L. J. (1993). Longitudinal study of the development of derived relations in an infant. *Journal of Experimental Child Psychology*, *56*(2), 201–239. https://doi.org/10.1006/jecp.1993.1032

May, R., Tyndall, I., McTiernan, A., Roderique-Davies, G., & Mcloughlin, S. (2022). The impact of the SMART program on cognitive and academic skills: a systematic review and meta-analysis. *British Journal of Educational Technology*. *53*, 1244 –1261. https://doi.org/10.1111/bjet.13192

McHugh, L., Barnes-Holmes, Y., & Barnes-Holmes, D. (2004). Perspective-taking as relational responding: A developmental profile. *Psychological Record*, *54*(1), 115–144. https://doi.org/10.1007/BF03395465

McKeel, A. N., Dixon, M. R., Daar, J. H., Rowsey, K. E., & Szekely, S. (2015). Evaluating the efficacy of the PEAK relational training system using a randomized controlled trial of children with autism. *Journal of Behavioral Education*, *24*(2), 230–241. https://doi.org/10.1007/s10864-015-9219-y

McLoughlin, S., Tyndall, I., & Pereira, A. (2020). Relational operant skills training increases standardized matrices scores in adolescents: A stratified active-controlled trial. *Journal of Behavioral Education*. https://doi.org/10.1007/s10864-020-09399-x

McLoughlin, S., Tyndall, I., Pereira, A., & Mulhern, T. (2021). Non-verbal IQ gains from relational operant training explain variance in educational attainment: An active-controlled feasibility study. *Journal of Cognitive Enhancement, 5*(1), 35–50. https://doi.org/10.1007/s41465-020-00187-z

Moran, L., Stewart, I., McElwee, J., & Ming, S. (2014). Relational ability and language performance in children with Autism Spectrum Disorders and typically developing children: A further test of the TARPA protocol. *Psychological Record*, *64*(2), 233–251. https://doi.org/10.1007/s40732-014-0032-0

O’Hora, D., Pelaez, M., & Barnes-Holmes, D. (2005). Derived relational responding and performance on verbal subtests of the WAIS-III. *Psychological Record*, *55*(1), 155–175. https://doi.org/10.1007/BF03395504

O’Hora, D., Peláez, M., Barnes-Holmes, D., Rae, G., Robinson, K., & Chaudhary, T. (2008). Temporal relations and intelligence: Correlating relational performance with performance on the WAIS-III. *The Psychological Record*, *58*(4), 569–583. https://doi.org/10.1007/bf03395638

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.r-project.org

Rusticus, S. A., & Lovato, C. Y. (2014). Impact of sample size and variability on the power and type I error rates of equivalence tests: A simulation study. *Practical Assessment, Research and Evaluation*, *19*(11), 1–10. https://doi.org/https://doi.org/10.7275/4s9m-4e81

Schmalz, X., Manresa, J. B., & Zhang, L. (2021). What is a Bayes factor? *Psychological Methods*. https://doi.org/10.1037/met0000421

Schönbrodt, F. D., Wagenmakers, E. J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339. https://doi.org/10.1037/met0000061

Slattery, B., & Stewart, I. (2014). Hierarchical classification as relational framing. *Journal of the Experimental Analysis of Behavior*, *101*(1), 61–75. https://doi.org/10.1002/jeab.63

Steele, D., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior*, *56*(3), 519–555.

Thirus, J., Starbrink, M., & Jansson, B. (2016). Relational Frame Theory, mathematical and logical skills: A multiple exemplar training intervention to enhance intellectual performance. *International Journal of Psychology and Psychological Therapy*, *16*(2), 141–155. http://urn.kb.se/resolve?urn=urn:nbn:se:miun:diva-28492

Vizcaíno-Torres, R. M., Ruiz, F. J., Luciano, C., López-López, J. C., Barbero-Rubio, A., & Gil, E. (2015). The effect of relational training on intelligence quotient: A case study. *Psicothema*, *27*(2), 120–127. https://doi.org/10.7334/psicothema2014.149

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686