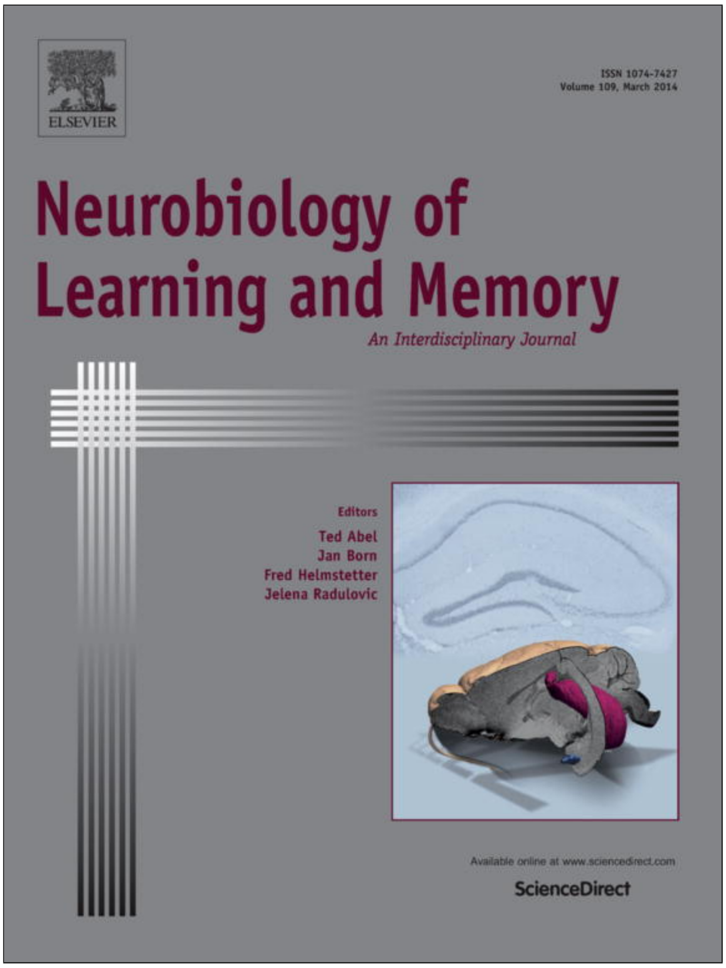
ELSEVIER
ISSN 1074-7427
Volume 109, March 2014
Neurobiology of Learning and Memory
An Interdisciplinary Journal
Editors
Ted Abel
Jan Born
Fred Helmstetter
Jelena Radulovic
Available online at www.sciencedirect.com
ScienceDirect

Neurobiology of Learning and Memory 109 (2014) 169–177

Contents lists available at ScienceDirect

Neurobiology of Learning and Memory

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

Relational memory generalization and integration in a transitive inference task with and without instructed awareness

Anita Munnelly [a,b,1], Simon Dymond [a,*]

[a] Department of Psychology, Swansea University, Singleton Park, Swansea SA2 8PP, United Kingdom
[b] School of Psychology, University College Dublin, Belfield, Dublin 4, Ireland

A R T I C L E I N F O

Article history:
Received 3 July 2013
Revised 6 January 2014
Accepted 8 January 2014
Available online 22 January 2014

Keywords:
Relational memory
Integration
Generalization
Hippocampus
Transitive inference
Awareness
Learning

A B S T R A C T

Two experiments investigated the potential facilitative effects of prior instructed awareness and predetermined learning criteria on humans' ability to make transitive inference (TI) judgments. Participants were first exposed to a learning phase and required to learn five premise pairs (A+B−, B+C−, C+D−, D+E−, E+F−). Testing followed, where participants made judgments on novel non-endpoint (BD, BE and CE) and endpoint inferential pairs (AC, AD, AE, AF, BF, CF and DF), as well as learned premise pairs. Across both experiments, one group were made aware that the stimuli could be arranged in a hierarchy, while another group were not given this instruction. Results demonstrated that prior instructional task awareness led to a minor performance advantage, but that this difference was not significant. Furthermore, in Experiment 2, inferential test trial accuracy was not correlated with a post-experimental measure of awareness. Thus, the current findings suggest that successful TI task performance may occur in the absence of awareness, and that repeated exposure to learning and test phases may allow weak inferential performances to emerge gradually. Further research and alternative methods of measuring awareness and its role in TI are needed.

1. Introduction

Considerable research has been conducted on the role of conscious awareness in learning and memory. Studies have found evidence for learning in the absence of awareness (e.g., Bayley, Frascino, & Squire, 2005; Cleeremans, Destrebecqz, & Boyer, 1998), while other findings indicate that awareness may (Clark & Squire, 1998; Cohen & Eichenbaum, 1993) or may not (e.g., Mudrik, Breska, Lamy, & Deouell, 2011; Purkis & Lipp, 2001; Williams, 2005) be needed for some forms of higher-order memory tasks requiring the generalization and integration of learning. One such widely studied task is the transitive inference (TI) task, which is considered a hallmark of human and nonhuman reasoning abilities (Vasconcelos, 2008). In a typical TI task, overlapping pairs of simultaneous discriminations are trained, such as A+B−, B+C−, C+D−, and D+E− (where "+" indicates reinforced choices and "−" non-reinforced choices), before novel combinations of stimulus pairs (e.g., AE, BD) are presented in a test phase in the absence of feedback. Now, the AE test pair may be solved without reference to the intervening pairs, and A chosen over E, since during training A is always reinforced and E is never reinforced. On the other hand, the BD pair have comparable training histories since reinforcement is made available equally often for choices of B and D during training. Despite this, selection of B over D is predicted because D was paired with E during training, which was never reinforced (Ellenbogen, Hu, Payne, Titone, & Walker, 2007; Frank, Rudy & O'Reilly, 2003; Frank, Rudy, Levy, & O'Reilly, 2005; Greene, Spellman, Dusek, Eichenbaum, & Levy, 2001; Lazareva & Wasserman, 2010; Libben & Titone, 2008; Martin & Alsop, 2004; Merritt & Terrace, 2011; Moses, Villate, Binns, Davidson, & Ryan, 2008; Moses, Villate, & Ryan, 2006; Werchan & Gomez, 2013).

The ability to generalize learning and integrate information in TI tasks may be mediated by conscious awareness (Libben & Titone, 2008; Martin & Alsop, 2004; Moses et al., 2006, 2008). "Awareness" in this instance is said to refer to a conscious understanding that the stimuli may be ordered along a hierarchy, a representation of which is then inspected and used to make inferential judgements (Greene et al., 2001; Smith & Squire, 2005). Task awareness is usually measured via post-experimental questionnaires, despite the acknowledged limitations of such measures (e.g., Lovibond & Shanks, 2002). Greene et al. (2001) instructed one group of participants that the stimuli formed a hierarchy, while another group were instructed to learn the pairs by trial

* Corresponding author. Fax: +44 1792 295679.
E-mail addresses: anita.munnelly@ucd.ie (A. Munnelly), s.o.dymond@swansea.ac.uk (S. Dymond).
[1] Fax: +44 1792 295679.

http://dx.doi.org/10.1016/j.nlm.2014.01.004
1074-7427/© 2014 Elsevier Inc. All rights reserved.

and error (see also, Libben & Titone, 2008). Measuring BD performance early during the learning phase assessed relational reasoning: after each block of learning trials, BD trials were tested (and retested). Greene et al. examined effects on subsequent performance and task awareness and found that instructed awareness led to faster learning and marginally greater accuracy on BD trials. Despite this, the group given no instructions performed significantly above chance on the BD test pair and that successful inferential performance was not correlated with a post-experimental measure of awareness. Furthermore, it was found that when performance on BD was near perfect, task awareness was not high, leading Greene et al. (2001) to propose that although explicit instructed awareness of the stimulus hierarchy is sufficient for TI, it may not be necessary. Additional, conflicting evidence however suggests that awareness of the stimulus hierarchy may in fact be a critical factor in TI (Martin & Alsop, 2004; Moses et al., 2006, 2008; Smith & Squire, 2005).

Evidence for the role of awareness, and whether or not instructed awareness facilitates generalization and integration of relational information, is therefore inconclusive and confounded by methodological shortcomings. For instance, the two studies to manipulate instructed awareness have employed differing learning and testing protocols and yielded contrasting results (Greene et al., 2001; Libben & Titone, 2008). That is, presenting the critical BD probe trials during learning may more readily identify the time course of the emergence of awareness and successful inferential performance, but it is confounded by the instructional manipulation. What are needed are unambiguous, separate demonstrations of the role of instructed awareness of the stimulus hierarchy and early BD testing on TI and post-experimental awareness measures. In addition, it may be beneficial to determine whether factors other than awareness, such as repeated exposure to learning and test phases enhance TI. That is, the ability to display accurate TI performances at test may rely on the establishment of stable patterns of responding during learning and test phases, rather than awareness.

Thus, in the present study we investigated the relationship between instructed awareness of the stimulus hierarchy, post-experimental awareness measures and methodological factors such as early BD probing and repeated learning and testing on the generalization and integration of memory in TI. In Experiment 1, one group of participants were instructed that the stimuli could be arranged in a hierarchy (Instructed group), while a second group were not given this information (Uninstructed group). Both groups then learned five premise pairs (A+B−, B+C−, C+D−, D+E− and E+F−), followed by testing with inferential, non-endpoint (BD, BE and CE), and endpoint pairs (AC, AD, AE, AF, BF, CF and DF) as well as the previously learned premise pairs (AB, BC, CD, DE and EF). Non-endpoint pairs represent the key trials in which generalized relational learning is demonstrated as all stimuli have equal histories of reinforcement and non-reinforcement (unlike endpoint pairs in which one stimulus is always reinforced and the other non-reinforced; Vasconcelos, 2008). In addition, Experiment 1 adopted a predetermined test mastery criterion and repeated exposure to learning and test phases, where necessary. Experiment 2 was similar except that the groups were presented with BD trials during learning trial-blocks and administered a post-experimental awareness questionnaire.

2. Materials and method: Experiment 1

2.1. Participants

Forty-three students, 20 men and 23 women, ranging in age from 18 to 33 years ($M_{age} = 20.51$, $SD = 2.59$) were recruited via the psychology subject pool at Swansea University. Participants were allocated partial course credit on completion of the study, and were randomly assigned to the Instructed ($n = 20$) or Uninstructed ($n = 23$) groups. Ethical approval was obtained from the Swansea University, Department of Psychology Ethics Committee prior to commencement.

2.2. Apparatus and stimuli

Six images randomly selected from the Kanji script were used as stimuli (see Fig. 1). The experimental task was programmed in E-Prime® (version 1.2), which controlled the presentation of all stimuli and recorded all responses.

2.3. Procedure

2.3.1. Learning phase

For participants in both groups, the learning phase began with the following onscreen instructions:

> During this phase you will be presented with two images in the middle right- and left-hand side of the computer screen. Your task is to learn to select the correct image. To select the image on the left, press the marked key on the left of the keyboard. To select the image on the right, press the marked key on the right of the keyboard. Sometimes the computer will give you feedback, and at other times it will not. The computer will tell you when this phase of the experiment is finished. Please press the spacebar to begin!

Additionally, participants in the Instructed group were told, *"There is an underlying hierarchy among the images. Your task is to learn this hierarchy"*; participants in the Uninstructed group were not given this information.

During this learning phase, all participants were presented with five adjacent stimulus pairs (A+B−, B+C−, C+D−, D+E− and E+F−; where "+" and "−" represent the reinforced and non-reinforced responses, respectively) and they were required to learn these pairs by trial and error. Both images from a pair were presented simultaneously in the middle of the computer screen. To select the

Fig. 1. The upper panel displays the Kanji images employed during training and testing. The images are labelled A, B, C, D, E and F (participants were never exposed to these labels). Also displayed are the premise, endpoint and non-endpoint test pairs.

image on the right, participants pressed the "m" key on the computer keyboard, while participants pressed the "z" key to select the image on the left. Both images remained onscreen until the participant made a selection. Left- and right- screen position was counterbalanced across trials. During learning, trials were followed by feedback presented in white on a black background. For example, when the premise pair AB was presented, selections of A were reinforced with the word "Correct!" while selections of B were unreinforced with the word "Wrong" (A+B−). Similarly, when the premise pair BC was presented, correct selections of B were reinforced, while incorrect selections of C were unreinforced (B+C−). The learning phase proceeded in this same manner for the remaining three pairs CD (C+D−), DE (D+E−) and EF (E+F−). Feedback remained onscreen for 1.5 s and was followed by an inter-trial interval (ITI) of 1.5 s. The five premise pairs were presented in a quasi-random order, four times each, within a block of twenty trials. In order to successfully complete this phase, participants were required to achieve 90% accuracy (i.e., make 18 out of 20 correct responses) on a given block. Learning blocks were repeated, if necessary, until this criterion was met.

2.3.2. Testing

On reaching criterion during the learning phase, participants proceeded immediately to the test phase, where all feedback was omitted. During this test phase, participants were presented with probes for the maintenance of the five premise pairs, alongside probes for seven endpoint and three non-endpoint pairs (see Table 1). Each test pair was presented four times, resulting in a total of sixty test trials. In order to meet criterion at testing, participants were required to achieve a minimum mean of 80% accuracy (i.e., 48 out of 60 correct responses), across all test pairs (premise, endpoint and non-endpoint). If participants failed to meet this criterion, they were re-exposed to the learning phase, followed again by testing for a maximum of three further times.

2.4. Statistical analysis

Separate two-way mixed between (Group: Instructed/Uninstructed) and within (Test Pair Type) analysis of variance (ANOVA) were conducted to examine accuracy on the premise, endpoint and non-endpoint test pairs. Post hoc comparisons were conducted using pairwise comparisons.

3. Results and discussion

Initially, forty-three participants began Experiment 1 but four from the Uninstructed group terminated their participation before the maximum four exposures to testing (Instructed $n = 20$ and Uninstructed $n = 19$, respectively). Four participants from the Instructed group and seven from the Uninstructed group failed to meet criterion during testing. Thus, 80% of Instructed participants (16/20) and 63% of Uninstructed participants (12/19) met the test criterion. Results demonstrated no significant differences in the number of participants in the Instructed and Uninstructed groups that met criterion at test (χ^2 (1, $n = 39$) = 1.46, $p = .243$). All participants from both groups were included in the analyses, regardless of whether or not they met the 80% test mastery criterion. Instructed participants were exposed to the test phase a mean of 2.10 times ($SD = 1.29$), and Uninstructed participants a mean of 1.9 times ($SD = 1.65$). No significant difference was found in the number of exposures to testing between groups ($t(38) = .001$, $p = 1.000$).

3.1. Learning trials to criterion

The Instructed group required between 20 and 680 trials ($M = 107.14$, $SD = 132.28$) and the Uninstructed group between 20 and 860 trials ($M = 116.86$, $SD = 197.11$) to meet the learning criterion. Groups did not differ on the mean number of trials required to reach criterion, $t(37) = -1.304$, $p = .20$ (see also Supplementary Material).

3.2. Accuracy: testing

3.2.1. Premise pairs

Fig. 2A displays the mean percentage correct on all premise pairs. A two-way mixed ANOVA with premise pair (AB, BC, CD, DE and EF) as within subjects factors and group (Instructed/Uninstructed) as between subjects factor was conducted. Mauchly's test indicated that the assumption of sphericity had been violated

Table 1
Test trials employed in Experiments 1 and 2.

	Test trial type						
Premise pairs	AB	BC	CD	DE	EF		
Endpoint pairs	AC	AD	AE	AF	BF	CF	DF
Non-endpoint pairs	BD	BE	CE				

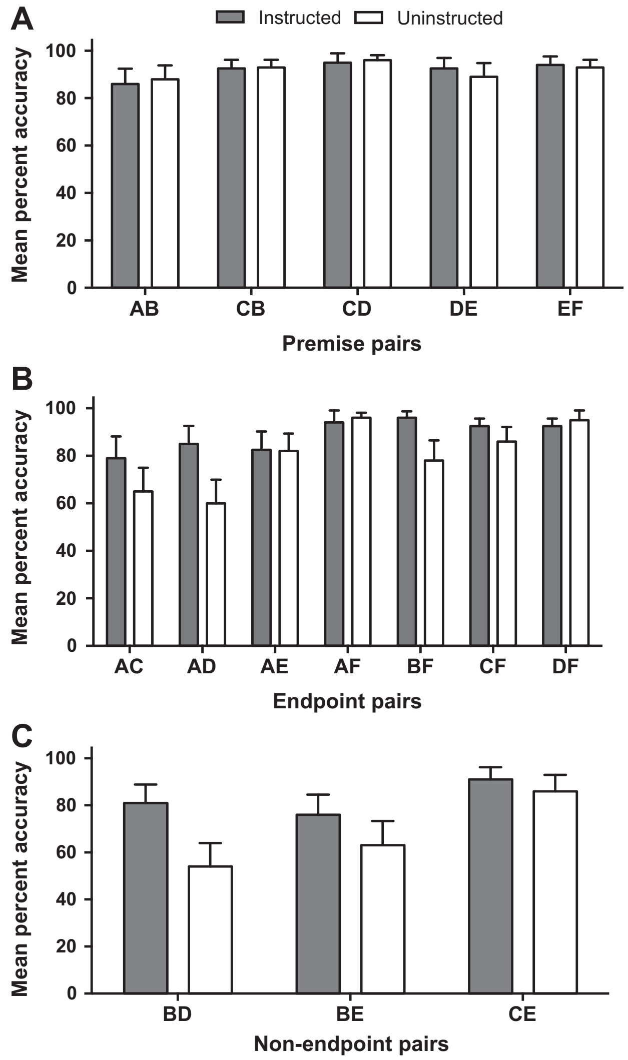

Fig. 2. Mean percentage accuracy on, (A) premise pairs, (B) endpoint pairs, and (C) non-endpoint pairs, for participants in the Instructed and Uninstructed groups in Experiment 1. Error bars indicate standard error.

($\chi^2(9) = 30.166$, $p = .001$), therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .807$; see Fields, 2013). This analysis revealed no main effect for premise pair ($F(4, 148) = 1.081$, $p = .368$; $\eta_p^2 = .028$) or group ($F(1, 37) = .001$, $p = .973$; $\eta_p^2 = .001$). In addition, there was no interaction between group and premise pair ($F(4, 148) = .100$, $p = .967$; $\eta_p^2 = .121$). This suggests that accuracy was similar on the premise pairs, and did not differ between the groups.

3.2.2. Endpoint pairs

Fig. 2B displays the mean percentage correct on all endpoint pairs. A two-way mixed ANOVA with endpoint pair (AC, AD, AE, AF, BF, CF and DF) as within subjects' factors and group (Instructed/Uninstructed) as between subjects factor was conducted. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(20) = 102.066$, $p = .001$), therefore degrees of freedom were corrected using Greenhouse Geisser estimates of sphericity ($\varepsilon = .553$). This analysis revealed a significant main effect for endpoint pair ($F(6, 222) = 5.086$, $p = .002$; $\eta_p^2 = .121$). Post hoc comparisons revealed that the accuracy was significantly higher on the endpoint pair, AF, over the endpoint pairs, AC and AD ($p = .048$). In addition, there was neither a main effect for group ($F(1, 37) = 2.109$, $p = .155$; $\eta_p^2 = .054$) nor was there an interaction between group and endpoint pair ($F(6, 222) = 1.699$, $p = .165$; $\eta_p^2 = .044$). Thus, accuracy did not differ on the endpoint pairs, or between the groups.

3.2.3. Non-endpoint pairs

Fig. 2C displays the mean percentage correct on all non-endpoint pairs. A two-way mixed ANOVA with non-endpoint pair (BD, BE and CE) as within subjects factors and group (Instructed/Uninstructed) as between subjects factor revealed a significant main effect for non-endpoint pair ($F(2, 74) = 6.666$, $p = .002$; $\eta_p^2 = .103$). Post hoc comparisons showed that accuracy was significantly higher on the non-endpoint pair CE over BD ($p = .012$). In addition, there was no main effect for group ($F(1, 37) = 2.722$, $p = .107$; $\eta_p^2 = .069$) and no interaction between group and non-endpoint pair ($F(2, 74) = 1.561$, $p = .217$; $\eta_p^2 = .040$). This shows that groups performed similarly on non-endpoint test pairs and that accuracy was higher on CE trials than on BD trials. In addition, for the Instructed group, accuracy on the BD non-endpoint pair was greater than chance (binomial probability test, $p < .05$), while for the Uninstructed group, accuracy was not greater than that expected by chance (binomial probability test, $p > .05$).

In summary, minimal evidence of between-group differences was found on accuracy to endpoint and non-endpoint pairs, as well as the premise pairs, with high accuracy on the non-endpoint pair CE over BD observed for both groups. Overall, findings from Experiment 1 indicate that while awareness of the stimulus hierarchy may have a facilitative effect on generalization and integration of relational memory in a TI task, it may not be necessary. Sixty-three percent of participants who were not given instructions also met the test criterion and required a similar number of trials to meet the criterion as those given instructions. In this way, prior instructed awareness of the underlying hierarchy may be necessary, but not sufficient for accurate relational learning to occur.

Our findings are similar to Greene et al. (2001), although it is important to note that we used a 6-term series TI task, whereas Greene et al. employed a 5-term series. It remains unclear whether or not the Uninstructed group in Greene et al. would have performed above chance levels on the additional non-endpoint pairs (BE and CE) presented in the current study. Furthermore, and in contrast to Experiment 1, Greene et al. exposed all participants to a test phase involving only the BD pair at the end of each learning block. For example, if participants made a minimum of seven out of eight correct responses on the BD test pair during the first block of testing, the experiment ended, and participants were required to complete the post-experimental awareness questionnaire. If participants did not meet this criterion, the experiment continued and awareness was only assessed on completion of the entire experiment (Greene et al., 2001). Therefore, it may be possible that presentation of BD pairs at the end of learning blocks in that study enhanced generalization and integration in the TI task for the Uninstructed group.

4. Experiment 2

In Experiment 2, we investigated whether presenting the inferential test pair, BD, throughout learning phases would lead to greater accuracy on other inferential trials (e.g., BE and CE) at test. A secondary aim was to examine association between performance and post-experimental measures of awareness (Greene et al., 2001; Lazareva & Wasserman, 2010).

5. Materials and method

5.1. Participants

Fifty-eight students, twenty-six men and thirty-two women, ranging in age from 18 to 49 years ($M_{age} = 21.61$, $SD = 4.92$) were recruited via the psychology subject pool at Swansea University. Participants were allocated partial course credit, or paid £6, on completion of the study, and were randomly assigned to the Instructed ($n = 29$) or Uninstructed ($n = 29$) group at the start of the experiment. Ethical approval was obtained from the Swansea University, Department of Psychology Ethics Committee prior to commencement.

5.2. Procedure

The procedure for Experiment 2 was identical to Experiment 1, with the exception that during the learning phase, participants were now presented with probe trials for the non-endpoint test pair, BD.

5.2.1. Learning phase

Similar to Experiment 1, participants were exposed to learning with five premise pairs (A+B−, B+C−, C+D−, D+E− and E+F−). However, in addition, four probe trials were presented, in which the BD test pair was presented, in the absence of feedback. That is, four BD probe trials were randomly interspersed among the premise pairs during learning blocks. Thus, participants were exposed to a total of twenty learning trials and four BD probe trials during the learning phase, and were required to achieve a minimum mean of 90% (i.e., 18 out of 20 correct responses) on the five premise pairs (AB, BC, CD, DE and EF). Accuracy on the BD probe pair did not affect mastery criterion during learning, and no other inferential (non-endpoint) probe trials were presented.

5.2.2. Testing

This phase was identical to Experiment 1.

5.2.3. Post-experimental awareness questionnaire

After completing the experiment, participants were provided with a questionnaire, which sought to determine what strategies, if any, were used to respond to novel test pairs. In order to determine an awareness score for each participant, two scorers were provided with a scoring guide for each question in the awareness questionnaire. Each question was assigned a score between 0 and 2 (where *2 = definite awareness*, *1 = some awareness*, and *0 = no awareness*; see also Moses et al., 2006). For instance, a score of 2 was awarded if a participant correctly ordered all six stimuli in the hierarchy, a score of 1 was awarded for two or less errors on this ordering, and a score of 0 was awarded if a participant made three or more errors on the ordering of the stimulus hierarchy (see Supplementary Materials for information on how each question was scored). Each participant received a total score out of 18 for their level of awareness. In order to calculate inter-observer agreement, the number of agreements between both scorers was divided by the number of agreements + disagreements × 100. Inter-observer agreement between both scorers was high (83%).

5.3. Statistical analysis

Separate two-way mixed between (Group: Instructed/Uninstructed) and within (Test Pair Type) Analysis of Variance (ANOVA) were conducted to examine accuracy on the premise, endpoint and non-endpoint test pairs. Post hoc comparisons were conducted using pairwise comparisons. In addition, Pearson correlations were conducted individually for the Instructed and Uninstructed groups on accuracy scores to the non-endpoint pairs, BD, BE and CE, using scores to Question 9 of the awareness questionnaire. Individual questionnaire scores for participants in the Instructed and Uninstructed groups can be seen in Tables 2 and 3.

6. Results and discussion

Initially, fifty-eight participants began Experiment 2 but four participants from the Uninstructed group terminated their participation before the maximum four exposures to testing (Instructed $n = 29$ and Uninstructed $n = 25$). Four participants from the Instructed group and five from the Uninstructed group failed to meet criterion during testing. Thus, 86% of Instructed participants (25/29) and 69% of Uninstructed participants (20/29) met the test criterion. Results demonstrated no significant differences in the number of participants in the Instructed and Uninstructed groups that met criterion at test (χ^2 (1, $n = 58$) = 2.48, $p = .115$). All participants from both groups were included in the analyses, regardless of whether or not they met the 80% test mastery criterion. Instructed and Uninstructed groups were exposed to the test phase a mean of 1.79 ($SD = 1.17$) and 2.32 ($SD = 1.15$) times, respectively, which did not differ statistically, ($t(52) = -1.584$, $p = .120$).

Table 2
Awareness scores to each question for participants in the Instructed group in Experiment 2.

Participant	Q.1	Q.2	Q.3	Q.4	Q.5	Q.6	Q.7	Q.8	Q.9	Total/18
Pass (n = 25)										
1	1	2	2	2	2	2	2	2	2	17
2	2	2	2	2	2	2	2	2	2	18
3	1	1	0	2	2	2	2	1	2	13
4	2	0	2	2	2	2	2	2	2	16
5	1	0	0	0	2	2	1	0	0	6
6	1	2	2	2	2	2	0	2	2	15
7	1	2	2	2	2	2	2	2	2	17
8	1	0	0	0	0	2	1	1	2	7
9	1	1	0	2	2	2	2	0	2	12
11	1	2	2	2	2	2	1	2	1	15
12	1	2	1	2	2	2	2	1	2	15
13	1	0	2	2	2	2	1	2	2	14
14	1	0	0	2	2	2	2	0	2	11
16	1	1	0	2	2	2	2	2	2	14
17	2	2	1	0	2	2	2	2	2	15
20	2	0	2	2	2	2	2	2	2	16
21	2	0	1	2	2	2	1	2	2	14
22	2	2	2	2	2	2	1	2	2	17
23	2	0	0	2	2	2	0	2	2	12
24	2	2	0	2	2	2	1	2	0	13
25	2	2	1	2	2	2	0	2	2	15
26	0	0	2	2	2	2	1	1	1	11
27	1	1	0	2	2	2	1	2	2	13
28	1	2	2	2	2	2	0	2	2	15
29	2	1	2	2	2	0	1	1	2	13
Fails (n = 4)										
10	1	2	0	2	0	2	2	0	2	11
15	1	2	2	0	0	2	2	0	2	11
18	2	2	2	0	2	2	2	0	2	14
19	2	0	0	0	0	0	1	0	2	5

Table 3
Awareness scores to each question for participants in the Uninstructed group in Experiment 2.

Participant	Q.1	Q.2	Q.3	Q.4	Q.5	Q.6	Q.7	Q.8	Q.9	Total/18
Pass (n = 20)										
1	2	1	0	2	2	2	2	2	1	14
3	1	0	0	2	2	2	1	2	2	12
8	1	0	2	2	2	2	2	1	2	14
9	2	1	0	0	2	2	2	2	2	13
11	0	2	2	2	2	2	0	2	2	14
12	1	0	1	0	0	2	1	0	0	5
13	1	0	2	2	2	2	2	2	0	13
15	1	1	0	2	2	2	2	0	2	12
16	1	2	2	2	2	2	2	2	2	17
17	2	1	1	2	0	2	1	1	0	10
18	1	2	2	0	0	2	2	0	2	11
19	1	0	2	2	2	2	2	0	2	13
20	2	0	2	0	2	2	1	0	2	11
21	1	2	1	2	2	2	0	2	2	14
22	0	0	2	2	2	2	0	2	2	12
23	2	0	1	0	0	2	1	0	2	8
24	1	2	2	0	0	2	0	0	2	9
25	2	1	1	2	2	2	1	2	2	15
28	2	0	1	2	2	2	0	2	2	13
29	2	1	0	2	2	2	1	2	0	12
Fails (n = 5)										
5	1	2	1	2	0	0	0	0	2	8
6	1	1	0	0	0	2	2	0	2	8
10	1	1	0	0	0	2	2	0	2	8
26	0	0	0	0	0	0	0	0	1	1
27	1	2	2	0	0	0	0	0	1	6

6.1. Learning trials to criterion

The Instructed group required between 24 and 384 trials ($M = 114.46$, $SD = 95.84$), while the Uninstructed group required between 24 and 432 trials ($M = 147.34$, $SD = 176.94$) to meet criterion. This difference was not significant, $t(52) = -.146$, $p = .883$; see also Supplementary Data).

6.2. Accuracy: testing

6.2.1. Premise pairs

Fig. 3A displays the mean percentage correct on all premise pairs. A two-way mixed ANOVA with premise pair (AB, BC, CD, DE and EF) as within subjects factors and group (Instructed/Uninstructed) as between subjects factor revealed no main effect for premise pair ($F(4, 208) = 1.245$, $p = .293$; $\eta_p^2 = .023$) or group ($F(1, 52) = .296$, $p = .588$; $\eta_p^2 = .006$), and nor was there an interaction between group and premise pair ($F(4, 208) = .933$ $p = .446$; $\eta_p^2 = .018$). As found in Experiment 1, this indicates that both groups performed similarly, and accuracy did not differ on the premise pairs at testing.

6.2.2. Endpoint pairs

Fig. 3B displays the mean percentage correct on all endpoint pairs. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(20) = 78.664$, $p = .001$), therefore degrees of freedom were corrected using Greenhouse Geisser estimates of sphericity ($\varepsilon = .608$). A two-way mixed ANOVA with endpoint pair

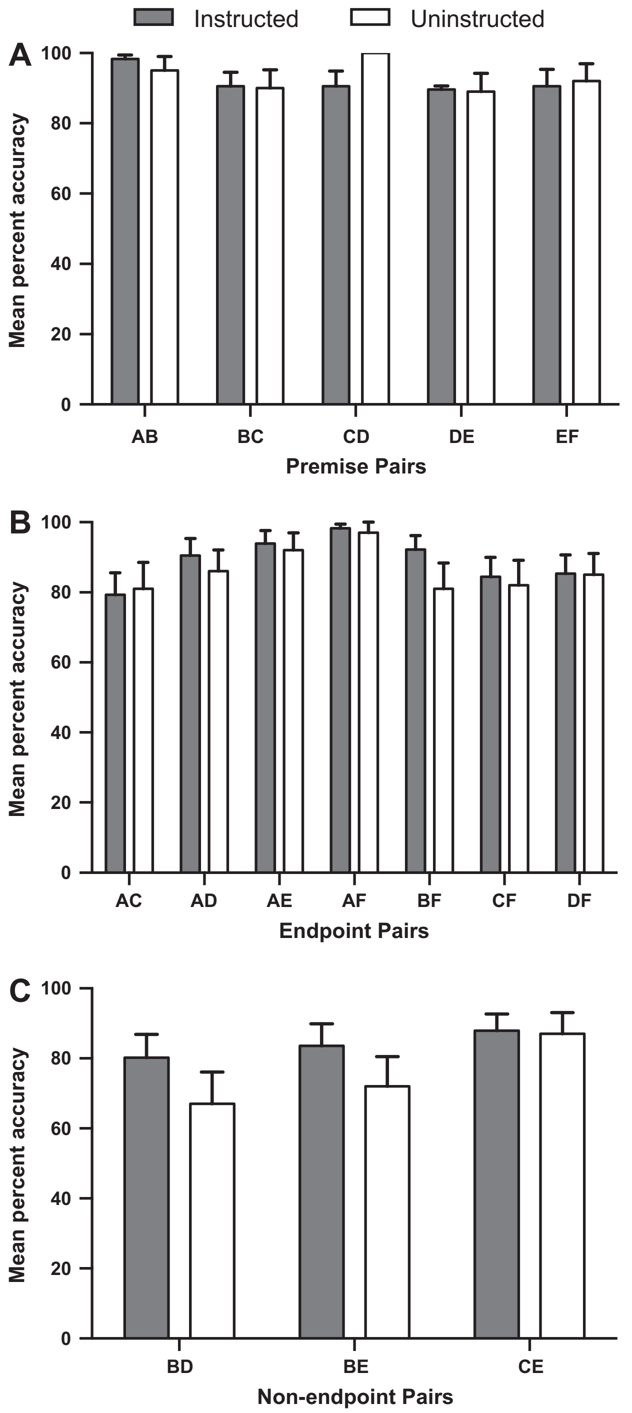

Fig. 3. (A) Mean percentage accuracy on, (A) premise pairs, (B) endpoint pairs, and (C) non-endpoint pairs, for participants in the Instructed and Uninstructed groups in Experiment 2. Error bars indicate standard error.

(AC, AD, AE, AF, BF, CF and DF) as within subjects factors and group (Instructed/Uninstructed) as between subjects factor revealed a main effect for endpoint pair ($F(6,\ 312) = 3.038$, $p = .022$; $\eta_p^2 = .055$). Post hoc comparisons revealed that accuracy was significantly higher on the endpoint pair AF, over the endpoint pair, AC ($p = .006$). No other differences were observed. In addition, there was no main effect for group ($F(1,\ 52) = .368$, $p = .814$; $\eta_p^2 = .009$) and no interaction between group and non-endpoint pair ($F(6, 312) = .368$, $p = .814$; $\eta_p^2 = .007$). In summary, accuracy differed on the endpoint pairs AF and AC, but not between the groups.

Table 4
Pearson correlation scores for awareness to Question 9 and test accuracy for participants in the Instructed and Uninstructed groups in Experiment 2 to the non-endpoint pairs, BD, BE and CE.

Pair	Instructed	Uninstructed
Non-endpoint pairs		
BD	.460	.524
BE	.098	.390
CE	−.181	−.788

6.2.3. Non-endpoint pairs

Fig. 3C displays the mean percentage correct on all non-endpoint pairs. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(2) = 9.136$, $p = .010$), therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .903$). A two-way mixed ANOVA with non-endpoint pair (BD, BE and CE) as within subjects factors and group (Instructed/Uninstructed) as between subjects factor revealed that a main effect for non-endpoint pair just missed significance ($F(2, 104) = 3.004$, $p = .060$; $\eta_p^2 = .055$). In addition, there was no main effect for group ($F(1, 52) = 1.342$, $p = .252$; $\eta_p^2 = .025$) and no interaction between group and non-endpoint pair ($F(2,\ 104) = .703$, $p = .484$; $\eta_p^2 = .013$). Thus, accuracy did not differ on the non-endpoint pairs, or between the groups. In addition, for both the Instructed and Uninstructed groups, accuracy on the BD non-endpoint pair was greater than chance (binomial probability test, $p < .05$).

7. Additional analysis

To determine whether the introduction of BD probe trials during the learning phase of Experiment 2 had a facilitative effect on performance accuracy on the non-endpoint pair BD, further analyses were conducted. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(2) = 6.740$, $p = .034$), therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .972$). A 2 (Group: Instructed/Uninstructed) × 2 (Experiment: 1: No BD probe trials; Experiment 2: BD probe trials) × 3 (Non-endpoint pair: BD, BE and CE) mixed ANOVA revealed a main effect for non-endpoint pair ($F(2, 180) = 9.919$, $p = .001$; $\eta_p^2 = .099$). Post hoc comparisons revealed that accuracy was significantly higher on the non-endpoint pair CE over BD ($p = .001$) and BE ($p = .001$). In addition, this analysis revealed no main effect for group ($F(1,\ 90) = 2.432$, $p = .122$; $\eta_p^2 = .026$), or experiment ($F(1,\ 90) = 1.493$, $p = .225$; $\eta_p^2 = .016$), nor was there an interaction between non-endpoint pair and experiment ($F(2,\ 180) = 1.096$, $p = .335$; $\eta_p^2 = .012$) or non-endpoint pair and group ($F(2,\ 180) = 1.143$, $p = .320$; $\eta_p^2 = .013$). Thus, neither the instructed awareness nor introduction of the BD trials during the learning phase in Experiment 2 influenced inferential accuracy at test.

8. Correlations between test accuracy and post-experimental measures of awareness

As mentioned, Pearson correlations were conducted individually for the Instructed and Uninstructed groups on accuracy scores to the non-endpoint pairs, BD, BE and CE, using scores to Question 9 of the awareness questionnaire. This analysis revealed no correlation between task awareness and accuracy on the BD, BE and CE non-endpoint pairs, for either the Instructed or Uninstructed groups (all p's > .05; see Table 4).

9. General discussion

In this study, we investigated the role of instructed task awareness in the generalization and integration of relational memory in a TI task, and sought to determine the potential facilitative effects of adopting a predetermined test mastery criterion and repeated exposures to learning and test phases. Results from Experiments 1 indicated that 80% of the Instructed group, and 63% of the Uninstructed group, met criterion at test. Similarly, in Experiment 2, 86% of the Instructed group and 69% of the Uninstructed group met test criterion. Thus, results from Experiments 1 and 2 demonstrated that awareness of the underlying stimulus hierarchy may not be necessary for successful task performance. Experiment 2 further revealed that awareness of the hierarchy was not correlated with accuracy on non-endpoint test pairs both for participants who did and did not receive the instructed awareness manipulation.

9.1. Role of instructed task awareness

In both experiments, accuracy on the endpoint and non-endpoint test pairs did not differ between the Instructed and Uninstructed groups. In addition, although minor improvements were noted for the Uninstructed group in Experiment 2 on non-endpoint pairs with the introduction of the BD probe trials during the learning phase, the difference in accuracy between Experiments 1 and 2 on the non-endpoint pairs was not significant. Thus, taken together, findings from Experiments 1 and 2 suggest that awareness of the stimulus hierarchy may not be necessary for successful TI task performance.

Experiment 2 of the current study sought to determine the potential facilitative effects associated with the introduction of BD probe trials during the learning phase, for the Uninstructed group. That is, would the introduction of BD probe trials during learning, facilitate accurate responding on the non-endpoint pairs at test? Indeed, previous research suggests that correct selections of B over D in the BD test pair is evidence for successful inferential responding, as both B and D are equally reinforced and non-reinforced during training (Vasconcelos, 2008). Furthermore, it has been proposed that awareness of the underlying stimulus hierarchy should be related to successful performance on the BD test pair (Smith & Squire, 2005). However, despite this, a number of studies have demonstrated successful inferential performance on the BD test pair in the absence of awareness (e.g., Frank, Rudy, Levy, & O'Reilly, 2005; Greene et al., 2001). For instance, a previous study by Greene et al. (2001) sought to determine whether presenting BD probe trials at the end of learning blocks, would more clearly determine the relationship between inferential performances at test, and awareness of the task. The authors found that the Uninstructed group performed significantly above chance on the BD test pair and that successful inferential performance was not correlated with a post-experimental measure of awareness. This finding led Greene et al. (2001) to propose that explicit instructed awareness of the stimulus hierarchy may not be necessary for TI. With respect to Experiment 2 of the current study, and in contrast to Greene et al., four unreinforced probe trials for the non-endpoint pair, BD, were presented *during* the learning phase. Although improvements in accuracy were noted for the Uninstructed group on the BD and BE non-endpoint pairs in Experiment 2 over those reported in Experiment 1, this difference was not significant. Furthermore, the Uninstructed group performed below-chance in Experiment 1 and above-chance in Experiment 2 on the BD non-endpoint pair. Although the reasons for the differences observed in accuracy of the Uninstructed group to the BD pair in the current study and those in the Greene et al. (2001) study remain unclear, procedural differences between the studies may at least partially account for the current findings. For instance, the Uninstructed group were exposed to a greater number of non-endpoint test pairs (e.g., BD, BE and CE), in comparison to Greene et al. Thus, it remains unclear whether the Uninstructed group in the Greene et al. study would have performed above chance on the BD non-endpoint pair if participants were exposed to additional non-endpoint test pairs (BE and CE), as in the current study. The inconsistencies between the current study and Greene et al. (2001), and the differences in accuracy on the BD non-endpoint pair between the studies when 5- and 6-items are employed, highlight the need for further research to determine the factors that influence the emergence of TI.

The finding that awareness of the stimulus hierarchy did not lead to significant differences in accuracy between the Instructed and Uninstructed groups on endpoint and non-endpoint test pairs also warrants further discussion. For example, considerable debate exists in the literature on TI as to whether awareness is necessary for individuals to respond to inferential problems at test. For instance, Frank et al. (2005) and Greene et al. (2001) found evidence for the expression of TI in adult humans in the absence of explicit awareness, whereas Moses et al. (2006) and Lazareva and Wasserman (2010) report that awareness of the stimulus hierarchy is necessary for successful inferential responding. With respect to Experiments 1 and 2 of the current study, results revealed that awareness of the stimulus hierarchy was not necessary for successful task performance, and that 63% (Experiment 1) and 69% (Experiment 2) of participants in the Uninstructed groups met criterion at test. In addition, Experiment 2 reported that responding on the inferential, non-endpoint pairs, BD, BE and CE, were not correlated with post-experimental measures of awareness, for either group. These findings are similar to Lazareva and Wasserman (2010), who reported that providing participants with additional instructions at the start of the experiment does not guarantee awareness at the end of the experiment. Taken with findings from Experiment 2 of the current study, these findings are important considering the debate regarding the role of awareness on the emergence of TI. Indeed, the disparity observed across studies regarding the role of awareness on TI, illustrates that alternative methods of examining the role of awareness on the emergence of TI might be needed. For example, the role of awareness may be more clearly identified by incorporating concurrent, trial-by-trial measures (Lovibond & Shanks, 2002). Concurrent probing of awareness may be considered preferable to the use of post-experimental questionnaires, which often follow a period of non-reinforced testing (Lovibond & Shanks, 2002), as was the case in Experiment 2, and permits the development of cognitive and behavioral processes to be tracked simultaneously during an experiment. Another alternative procedure involves identifying the role of private verbal behavior in operant learning by requiring participants to "talk aloud" throughout all experimental tasks (Cabello, Luciano, Gomez, & Barnes-Holmes, 2004; Hayes, 1986). Verbalizations may then be transcribed and correlations tested between the content of verbal categories and behavioral performance measures. These procedures aim to determine whether verbal behavior affects participants' ability to contact the programmed contingencies during learning and generalization tasks, by measuring both concurrently, and may prove useful in future studies seeking to examine the role of awareness in TI.

9.2. Role of learning and test mastery criterion and repeated test exposures

A secondary aim of the present study was to examine the role of learning and test mastery criterion and repeated test exposures on TI test performance. Previous studies have tended to present a fixed number of learning trial blocks and employ a lower accuracy criterion (e.g., 75% in Ellenbogen et al., 2007) on some but not all premise pairs, or have repeated trial blocks and presented fewer training trials in additional blocks (e.g., Heckers, Zalesak, Weiss, Ditman, & Titone, 2004; Smith & Squire, 2005). To our knowledge, the present study is the first to employ combined accuracy and exposure criteria during both learning and testing within a traditional TI paradigm. Previous neurobehavioral research with a 5-term hierarchy has employed similar learning and testing criteria but with contextually controlled matching-to-sample procedures examining arbitrary relations of more than and less than rather than the overlapping simultaneous discrimination format of the current experiments (e.g., Hinton, Dymond, von Hecker, & Evans, 2010; Munnelly, Dymond, & Hinton, 2010). In behavioral research on relational memory ability, the practices of ensuring premise pairs are learned to high accuracy and of incorporating combined mastery and exposure criteria during test phases is widely used to determine whether or not stable inferential performance has emerged (Dymond & Rehfeldt, 2000; Dymond & Roche, 2013). Methodological approaches like this are based on a functional view of learning (Chiesa, 1994; De Houwer, Barnes-Holmes, & Moors, 2013) in which behavior changes as a result of regularities in the environment. According to this approach, prior learning experiences with a limited subset of information (e.g., the premise pairs) are usually sufficient for relational integration and generalization to occur among untested, nonadjacent events (e.g., endpoint and non-endpoint pairs). Learning the premise pairs to high accuracy, *ab initio*, forms the basis on which novel relational memory generalization occurs. Because the expression of relational memory often emerges gradually and at a different pace across individuals, using a predetermined exposure criterion thus allows for both predicted and unpredicted performances to emerge. In this way, a functional approach makes no assumptions about the ability of individuals who failed tests for TI to subsequently pass such tests and only requires that their learning histories be relatively well matched and controlled prior to the crucial test phases. Moreover, it assumes that, for largely unspecified reasons, the same procedures may not be uniformly effective in leading to the expression of relational skills for all those tested. It becomes then an empirical issue to design tasks that maximize these environmental regularities and lead to the expression of complex relational skills such as TI in the majority of participants. The present approach with its emphasis on predetermined learning and exposure criteria and parsimony may therefore provide useful in further research on the cognitive, relational processes underlying TI (De Houwer, 2011; De Houwer et al., 2013).

9.3. Clinical neuroscience implications

The present findings may have implications for understanding the emergence of relational memory abilities in individuals that suffer from cognitive impairments, such as amnesia (e.g., Smith & Squire, 2005) and schizophrenia (e.g., Armstrong, Kose, Williams, Woolard, & Heckers, 2010; Coleman et al., 2010; Titone, Ditman, Holzman, Eichenbaum, & Levy, 2004). A number of studies have shown that TI is impaired in individuals with schizophrenia (e.g., Titone et al., 2004). More specifically, while Titone et al. found that BD performance was impaired in schizophrenia patients (see also, Coleman et al., 2010), this group was able to learn the premise pairs, and display high levels of accuracy on the endpoint pair AE, at comparable levels to controls. Interestingly, awareness of the stimulus hierarchy is unlikely to be a factor affecting the emergence of TI in schizophrenia because conscious awareness is not significantly associated with the ability to make inferential judgments in this group (Coleman et al., 2010). Armstrong, Williams, and Heckers (2012) have proposed that in fact individuals with schizophrenia suffer from a differential relational memory deficit and that task awareness is not the contributing factor in explaining TI impairment. In their study, the authors investigated the potential facilitative effects of exposing patients to reduced-sized learning blocks and additional feedback during learning phases in an attempt to maximize the number of patients exposed to the inferential tests. Only 8% (3/37) of the schizophrenia patients failed to complete the learning phase, which was an improvement on previous findings (Armstrong et al., 2010). The learning and testing protocol employed in the present experiments may therefore provide an alternative method to examine the emergence of TI in individuals suffering from cognitive impairments, such as schizophrenia. Furthermore, if additional feedback and reduced learning blocks were incorporated alongside additional learning and test phases, it may be possible to generate more accurate inferential responding in patients with schizophrenia. Clearly, these issues warrant further investigation.

Neuroimaging research has shown that TI performance in both healthy controls and clinical groups, such as those with schizophrenia and psychosis, involves differential activation of an extended network encompassing frontal and parietal lobes, hippocampus, midbrain (ventral tegmental area/substantia nigra), and anterior cingulate cortex (ACC) (Acuna, Eliassen, Donoghue, & Sanes, 2002; Heckers et al., 2004; Hinton et al., 2010; Ongur et al., 2006; Shohamy & Wagner, 2008; Van Elzakker, O'Reilly, & Rudy, 2003; Williams, Avery, Woolard, & Heckers, 2012). For instance, Shohamy and Wagner (2008) found that hippocampal activation did not differ between correct training trials and generalization trials, but that the magnitude of activation increase in the hippocampus and midbrain during learning was associated with subsequent accuracy on the generalization test. The correlation between hippocampal and midbrain activation during learning and subsequent test performance was not influenced by either the strength of retention of the trained pairs or by self-reported awareness of the stimulus hierarchy. Shohamy and Wagner's findings indicating that post-experimental awareness of the underlying hierarchy was largely independent of both task performance and midbrain/hippocampal modulation are consistent with the present findings from Experiment 2 showing that awareness was not correlated with test accuracy in either group. Instructed task learning, which resembles the present task awareness manipulation, is known to activate lateral prefrontal and parietal cortices (Cole, Laurent, & Stocco, 2013; Stocco, Lebiere, O'Reilly, & Anderson, 2012), although it remains to be seen whether or not these regions are recruited in an instructed TI task. Future research might, therefore, investigate the role of task awareness instructions on the differential recruitment of hippocampus, midbrain, frontal regions and ACC during relational learning and generalization testing.

10. Conclusion

The present findings were obtained with a relatively small sample size and suggest that task awareness and BD probe presentation could play a role in facilitating the generalization and integration of relational memory in a TI task. Further research with larger samples sizes is necessary to replicate the findings and determine the size of any possible effects (Lakens, 2013), if any, of task awareness and emergent probe presentation on TI abilities. Moreover, the potential utility of incorporating mastery criterion during test phases and of exposing participants to additional learning and test phases, if the predicted patterns of performance do not emerge initially, should be considered in future research on TI.

Overall, we found that awareness of the stimulus hierarchy may not be necessary for successful task performance. Experiment 2 revealed that accuracy on inferential test trials (BD, BE and CE) did not correlate with awareness. Neurobehavioral research is needed that incorporates alternative methods to determine the specific role of task awareness and the neural circuitry underpinning relational memory integration and generalization.

Acknowledgment

We thank Richard Moore for assistance with data collection.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.nlm.2014.01.004.

References

Acuna, B. D., Eliassen, J. C., Donoghue, J. P., & Sanes, J. N. (2002). Frontal and parietal lobe activation during transitive inference in humans. *Cerebral Cortex, 12*, 1312–1321.
Armstrong, K., Kose, S., Williams, L., Woolard, A., & Heckers, S. (2010). Impaired associative inference in patients with schizophrenia. *Schizophrenia Bulletin, 37*, 1–8.
Armstrong, K., Williams, L. E., & Heckers, S. (2012). Revised associative inference paradigm confirms relational memory impairment. *Neuropsychology, 26*(4), 451–458.
Bayley, P. J., Frascino, J. C., & Squire, L. R. (2005). Robust habit learning in the absence of awareness and independent of the medial temporal lobe. *Nature, 436*, 550–553.
Cabello, F., Luciano, C., Gomez, I., & Barnes-Holmes, D. (2004). Human schedule performance, protocol analysis, and the "silent dog" methodology. *The Psychological Record, 54*, 405–422.
Chiesa, M. (1994). *Radical behaviorism: The philosophy and the science*. Boston, MA: Authors' Cooperative.
Clark, R. E., & Squire, L. R. (1998). Classical conditioning and brain systems: The role of awareness. *Science, 280*, 77–81.
Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Science, 2*, 406–416.
Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia and the hippocampal system*. Cambridge, MA: MIT Press.
Cole, M. W., Laurent, P., & Stocco, A. (2013). Rapid instructed task learning: A new window into the human brain's unique capacity for flexible cognitive control. *Cognitive, Affective and Behavioral Neuroscience, 13*, 1–22.
Coleman, M. J., Titone, D., Krastoshevsky, O., Krause, V., Haans, Z., Mendell, N. R., et al. (2010). Reinforcement ambiguity and novelty do not account for transitive inference deficits. *Schizophrenia Bulletin, 36*(6), 1187–1200.
De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science, 6*, 202–209.
De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin and Review, 20*, 631–642.
Dymond, S., & Rehfeldt, R. (2000). Understanding complex behavior: The transformation of stimulus functions. *The Behavior Analyst, 23*, 239–254.
Dymond, S., & Roche, B. (2013). *Advances in relational frame theory: Research & application*. Oakland, CA: New Harbinger.
Ellenbogen, J. M., Hu, P. T., Payne, J. D., Titone, D., & Walker, M. P. (2007). Human relational memory requires time and sleep. *Proceedings of the National Academy of Sciences of the United States of America, 18*, 7723–7728.
Fields, A. (2013). *Discovering statistics using SPSS* (4th ed.). London: Sage Publications Ltd..
Frank, M. J., Rudy, J. W., Levy, W. B., & O'Reilly, R. C. (2005). When logic fails: Implicit transitive inference in humans. *Memory and Cognition, 33*, 742–750.
Frank, M. J., Rudy, J. W., & O'Reilly, R. C. (2003). Transitivity, flexibility, conjunctive representations, and the hippocampus. II. A computational analysis. *Hippocampus, 13*, 341–354.
Greene, A. J., Spellman, B. A., Dusek, J. A., Eichenbaum, H. B., & Levy, W. B. (2001). Relational learning with and without awareness: Transitive inference using nonverbal stimuli in humans. *Memory and Cognition, 29*, 893–902.
Hayes, S. C. (1986). The case of the silent dog – Verbal reports and the analysis of rules: A review of Ericsson and Simon's Protocol analysis: Verbal reports as data. *Journal of the Experimental Analysis of Behavior, 45*(3), 351–363.
Heckers, S., Zalesak, M., Weiss, A. P., Ditman, T., & Titone, D. (2004). Hippocampal activation during transitive inference in humans. *Hippocampus, 14*, 153–162.
Hinton, E. C., Dymond, S., von Hecker, U., & Evans, C. J. (2010). Neural correlates of relational reasoning and the symbolic distance effect: Involvement of parietal cortex. *Neuroscience, 168*, 138–148. Corrigendum, 170, 1345.
Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Front. Psychol., 4*, 863. http://dx.doi.org/10.3389/fpsyg.2013.00863.
Lazareva, O. F., & Wasserman, E. A. (2010). Nonverbal transitive inference: Effects of task and awareness on performance. *Behavioural Processes, 83*, 99–112.
Libben, M., & Titone, D. (2008). The role of awareness and working memory in human transitive inference. *Behavioural Processes, 77*, 43–54.
Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes, 28*, 3–26.
Martin, N., & Alsop, B. (2004). Transitive inference and awareness in humans. *Behavioural Processes, 67*, 157–165.
Merritt, P. J., & Terrace, H. S. (2011). Mechanisms of inferential order judgments in humans (Homo sapiens) and rhesus monkeys (Macaca mulanta). *Journal of Comparative Psychology, 125*, 227–238.
Moses, S. N., Villate, C., Binns, M. A., Davidson, P. S. R., & Ryan, J. D. (2008). Cognitive integrity predicts transitive inference performance bias and success. *Neuropsychologia, 46*, 1314–1325.
Moses, S. N., Villate, C., & Ryan, J. D. (2006). An investigation of learning strategy supporting transitive inference performance in humans compared to other species. *Neuropsychologia, 44*, 1370–1387.
Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration without awareness: Expanding the limits of unconscious processing. *Psychological Science, 22*(6), 764–770.
Munnelly, A., Dymond, S., & Hinton, E. C. (2010). Relational reasoning with derived comparative relations: A novel model of transitive inference. *Behavioural Processes, 85*, 8–17.
Ongur, D., Cullen, T. J., Wolf, D. H., Rohan, M., Barreira, P., Zalesak, M., et al. (2006). The neural basis of relational memory deficits in schizophrenia. *Archives of General Psychiatry, 63*, 356–365.
Purkis, H. M., & Lipp, O. V. (2001). Does affective learning exist in the absence of contingency awareness? *Learning and Motivation, 32*(1), 84–99.
Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron, 60*, 378–389.
Smith, C., & Squire, L. R. (2005). Declarative memory, awareness, and transitive inference. *Journal of Neuroscience, 25*, 10138–10146.
Stocco, A., Lebiere, C., O'Reilly, R. C., & Anderson, J. R. (2012). Distinct contributions of the caudate nucleus, rostral prefrontal cortex, and parietal cortex to the execution of instructed tasks. *Cognitive, Affective and Behavioral Neuroscience, 12*, 611–628.
Titone, D., Ditman, T., Holzman, P. S., Eichenbaum, H., & Levy, D. L. (2004). A transitive inference test of relational memory in schizophrenia. *Schizophrenia Research, 68*, 235–247.
Van Elzakker, M., O'Reilly, R. C., & Rudy, J. W. (2003). Transitivity, flexibility, conjunctive representations and the hippocampus. I. An empirical analysis. *Hippocampus, 13*, 334–340.
Vasconcelos, M. (2008). Transitive inference in non-human animals: An empirical and theoretical analysis. *Behavioural Processes, 78*, 313–334.
Werchan, D. W., & Gomez, R. L. (2013). Generalizing memories over time: Sleep and reinforcement facilitate transitive inference. *Neurobiology of Learning and Memory, 100*, 70–76.
Williams, J. N. (2005). Learning without awareness. *Studies in Secondary Language Acquisition, 27*, 209–304.
Williams, L. E., Avery, S. N., Woolard, A. A., & Heckers, S. (2012). Intact relational memory and normal hippocampal structure in the early stage of psychosis. *Biological Psychiatry, 71*, 105–113.