PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Research

Cite this article: Ponari M, Norbury CF, Rotaru A, Lenci A, Vigliocco G. 2018 Learning abstract words and concepts: insights from developmental language disorder. *Phil. Trans. R. Soc. B* **373**: 20170140. http://dx.doi.org/10.1098/rstb.2017.0140

Accepted: 4 May 2018

One contribution of 23 to a theme issue 'Varieties of abstract concepts: development, use and representation in the brain'.

Subject Areas:
behaviour, cognition

Keywords:
developmental language disorder, language acquisition, abstract concepts, distributional semantics, semantic representation, vocabulary development

Author for correspondence:
Gabriella Vigliocco
e-mail: g.vigliocco@ucl.ac.uk

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4097402.

THE ROYAL SOCIETY PUBLISHING

Learning abstract words and concepts: insights from developmental language disorder

Marta Ponari[1], Courtenay Frazier Norbury[2], Armand Rotaru[3], Alessandro Lenci[4] and Gabriella Vigliocco[3]

[1]School of Psychology, University of Kent, Canterbury, UK
[2]Language and Cognition, Psychology and Language Sciences, University College London, Chandler House, 2 Wakefield Street, London WC1N 1PF, UK
[3]Institute for Multimodal Communication, Psychology and Language Sciences, University College London, 26 Bedford Way, London WC1H 0AP, UK
[4]Computational Linguistics Laboratory, Department of Philology, Literature and Linguistics, University of Pisa, Pisa, Italy

GV, 0000-0002-7190-3659

Some explanations of abstract word learning suggest that these words are learnt primarily from the linguistic input, using statistical co-occurrences of words in language, whereas concrete words can also rely on non-linguistic, experiential information. According to this hypothesis, we expect that, if the learner is not able to fully exploit the information in the linguistic input, abstract words should be affected more than concrete ones. Embodied approaches instead argue that both abstract and concrete words can rely on experiential information and, therefore, there might not be any linguistic primacy. Here, we test the role of linguistic input in the development of abstract knowledge with children with developmental language disorder (DLD) and typically developing children aged 8–13. We show that DLD children, who by definition have impoverished language, do not show a disproportionate impairment for abstract words in lexical decision and definition tasks. These results indicate that linguistic information does not have a primary role in the learning of abstract concepts and words; rather, it would play a significant role in semantic development across all domains of knowledge.

This article is part of the theme issue 'Varieties of abstract concepts: development, use and representation in the brain'.

1. Introduction

Learning the meaning of words is one of the most complex and remarkable of human achievements. Learning words is hard because even when the referent is present in the physical environment, rarely is it isolated in the visual scene [1]. To make the situation worse, referents are not always present in the physical environment, either because they are spatially and/or temporally displaced (e.g. talk about past or future events), or because they are abstract and have no tangible referent.

A number of theories argue that abstract concepts are grounded (solely or primarily) in our *linguistic* experience [2–5], whereas concrete words could benefit also from non-linguistic information. For example, it has been shown that the richness of featural representations (used as a proxy of sensory-motor and affective content) predicts behavioural effects (e.g. lexical decision, semantic priming) better for concrete than abstract words, whereas the richness of the linguistic contexts in which a word appears (semantic neighbourhood density, used as a proxy for language-based information) predicts behavioural effects better for abstract than concrete words [6].

Embodied theories of semantic representation instead argue that learning and representing both concrete and abstract concepts are grounded in our

Table 1. Demographic characteristics of DLD and TD children and performance on the background tests, means (s.d.).

		age-matched			vocabulary-matched		
	DLD	TD_{age}	*t*-test	*p*-value	TD_{voc}	*t*-test	*p*-value
age	10.40 (1.83)	10.33 (1.44)	0.127	0.899	8.16 (2.12)	3.383	0.002
matrix reasoning	40.33 (10.67)	49.22 (9.37)	2.656	0.012	51.41 (8.31)	3.413	0.002
BPVS	108.72 (25.03)	129.66 (14.74)	3.059	0.004	109 (24.25)	0.034	0.973
CELF recall sentence	4.83 (4.23)	n.a.	—	—	n.a.	—	—

experience of the world. There is now plenty of evidence that processing concrete concepts in adults engages to some extent the same cognitive and neural systems involved in perceiving and acting upon the physical world [7]. There is also growing evidence that processing abstract concepts in adults involves motor representations [8,9], simulation of specific situations [10] and the emotion system [11,12]. In development, Ponari *et al.* [13] showed that abstract words with emotional connotations are learnt earlier than neutral abstract words, suggesting that emotion could serve as a bootstrapping mechanism for the learning of abstract words and concepts. Scholars who argue for a role of embodied information in the learning and representation of abstract concepts also assume that linguistic information matters but do not claim 'language primacy' [8,14,15].

Here, we present a test of the role of linguistic information in learning semantic representations for abstract words comparing the knowledge of abstract and concrete words by children with developmental language disorder (DLD) and their typically developing peers (TD).

DLD is a neurodevelopmental disorder affecting approximately 7.5% of children at school entry [16]. Children with DLD typically present with severe deficits in morphosyntax and other aspects of grammar [17] as well as vocabulary that is reduced in both breadth and depth relative to TD peers [18]. Vocabulary reduction in children with DLD has been linked to a number of different causes, among which are working memory deficits [19], statistical learning [20] and attention [21]. However, no previous study to our knowledge has focused on abstract words, despite the anecdotal report by speech and language professionals that these children are especially impaired as regards these words. Here, we investigate knowledge of abstract and concrete word meanings in children with DLD and TD peers matched for chronological age (TD_{age}) or receptive vocabulary scores (TD_{voc}). As DLD is assumed to affect vocabulary development [18], it follows that, if learning abstract words is based primarily on linguistic information, then abstract words should be disproportionately impaired relative to concrete words in children with DLD when compared with their TD peers. The inclusion of both age- and vocabulary-matched control groups allows us to assess both quantitative and qualitative differences in knowledge of words: the comparison with age-matched TD children can tell us whether DLD children show any quantitative difference from their peers. Thus, if DLD children show greater impairment for abstract than concrete words, this could be either because these words are learnt later by DLD children, or because there are qualitative differences in the manner in which DLD and TD children learn vocabulary. The comparison with younger vocabulary-matched TD children will then allow us to make inferences about whether any difference we find in the DLD–TD_{age} comparison depends on qualitative differences in the way DLD children use and organize their word knowledge, or whether DLD children are simply behind in their linguistic development.

We chose to use both definitions and lexical decision tasks: defining words provides a direct window into what children know about concepts; it is, however, a challenging task as it further requires expressive language, which is often compromised in children with DLD. Thus, the definition task may underestimate word knowledge in this group. The lexical decision does not require language production although it provides a more indirect window into children's knowledge of the word.

2. Methods

(a) Participants

Eighteen children with DLD (14 males; mean age = 10.03, s.d. = 1.76) were recruited from schools in southeast England. All children had a clinical diagnosis from a speech–language therapist external to the research team. Children in the control groups were selected from a pool of 73 TD children who completed both tasks: 18 children (14 males; mean age = 10.34, s.d. = 1.44) were matched to the DLD children on gender and age (TD_{age}), and 18 (14 males; mean age = 8.16, s.d. = 2.12) were matched to the DLD children on gender and raw scores on the British Picture Vocabulary Scale (BPVS, [22]; TD_{voc}). TD children were recruited from local schools and did not have any reported special educational needs, or history of language delay. Non-verbal cognitive abilities were assessed using the Matrix Reasoning test of the Wechsler Abbreviated Scale of Intelligence [23]. DLD children were also administered the Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals: Core Language Scales (CELF; [24]). The children's characteristics are summarized in table 1. The protocol was approved by the Research Ethics Committee at University College London; informed, written consent was obtained from all parents and verbal assent was obtained from all children prior to assessment. The same children participated in both tasks.

Table 2. Lexical and sublexical characteristics of the words used, means (s.d.).

variable	concreteness category: abstract	concreteness category: concrete	*t*-test	*p*-value
AoA band 1				
concreteness[a]	337.22 (45.7)	576.44 (35.8)	12.338	<0.001
length (no. of letters)	5.44 (1.0)	5.44 (1.0)	0.000	1.000
valence[b]	5.10 (1.9)	5.10 (1.9)	0.006	0.995
age of acquisition[c]	5.22 (0.94)	4.80 (0.80)	1.401	0.180
CBBC frequency[d]	4.7 (0.43)	4.64 (0.31)	0.505	0.621
familiarity[a]	566.33 (30.74)	565.33 (27.47)	0.049	0.962
AoA band 2				
concreteness[a]	319.56 (50.55)	509.22 (70.78)	6.542	<0.001
length (no. of letters)	4.67 (0.5)	4.67 (0.5)	0.000	1.000
valence[b]	4.93 (1.69)	4.90 (1.66)	0.025	0.980
age of acquisition[c]	7.16 (1.23)	6.67 (1.53)	0.756	0.460
CBBC frequency[d]	4.45 (0.45)	4.39 (0.40)	0.280	0.783
familiarity[a]	543.0 (16.18)	533.67 (23.02)	0.995	0.335
AoA band 3				
concreteness	334.11 (26.7)	517.78 (71.0)	7.265	<0.001
length (no. of letters)	5.67 (1.32)	5.78 (1.20)	0.000	1.000
valence[b]	4.94 (1.27)	4.75 (1.81)	0.28	0.807
age of acquisition[c]	9.04 (1.44)	9.14 (1.53)	0.157	0.877
CBBC frequency[d]	3.56 (1.32)	3.56 (120)	0.005	0.996
familiarity[a]	464.67 (61.01)	463.56 (74.72)	0.035	0.973
AoA band 4				
concreteness[a]	322.78 (41.37)	495.13 (67.38)	6.442	<0.001
length (no. of letters)	6.33 (1.32)	6.22 (1.09)	0.194	0.848
valence[b]	4.90 (1.47)	5.08 (1.42)	0.272	0.789
age of acquisition[c]	10.71 (0.78)	10.74 (0.47)	0.079	0.938
CBBC frequency[d]	3.26 (0.74)	3.05 (0.52)	0.680	0.506
familiarity[a]	430.56 (60.0)	448.38 (56.19)	0.630	0.538

(*Continued*.)

Table 2. (*Continued*.)

[a][30].
[b][27].
[c][25].
[d][29].

(b) Materials

Thirty-six abstract and 36 concrete words were selected from a pool of 3505 words for which normative data on a range of lexical variables could be obtained. These variables included: the age of acquisition (AoA; [25]), concreteness [26], valence [27] and log-frequency [28]. AoA ratings were used to ensure the items selected were appropriate for our participants' ages: words were divided into age of acquisition bands (1: words acquired at 4–5 years; 2: 6–7 years; 3: 8–9 years; 4: 10–11 years). Within each AoA band, concrete and abstract words were matched for valence, length (number of letters) and log-frequency. Concrete and abstract words also did not differ on familiarity, and on a measure of frequency taken from subtitles from a UK TV channel targeted at children aged 6–12 (CBBC; [29]). Lexical and sublexical characteristics of the words are listed in table 2; see the electronic supplementary materials for a list of all words and the non-words used in the lexical decision task.

Among these 72 words, 24 (12 abstract and 12 concrete) were shared between the two tasks; 24 (12 abstract and 12 concrete) were used for the definitions task only, and the remaining 24 were used for the auditory lexical decision task only. Additionally, for the lexical decision task, 48 pronounceable non-words were created by changing one phoneme from 48 words matched to the experimental words on length, AoA, valence and concreteness. All words and non-words were recorded by a native English speaker using Audacity v. 1.2.2 [31].

(c) Procedure

All children were assessed in their school and received stickers for participation. Stimuli were presented verbally using E-Prime v. 2.0 [32] running on a laptop with a touchscreen display. Participants were presented with short computer games in which they were asked to help a cartoon alien learn English. The lexical decision task was always presented before the definition task, in a single session. Children received verbal instructions from the experimenter, and were asked to wear headphones prior to the beginning of each task.

(i) Lexical decision

In each trial, a cartoon alien was presented in the middle of the screen for 1000 ms, followed by the auditory presentation of either a real English word or a non-word. Immediately after the offset of the word (average stimulus duration = 830 ms), two touch screen buttons appeared at the bottom left (a red thumbs-down icon) or the bottom right (a green thumbs-up icon) of the screen and children were asked to indicate whether what they heard was a word they knew (green button), or a 'funny, made-up' word (red button). Six practice trials (three non-words and three words not used in the experiment) included visual feedback of either a smiling (correct trial) or frowning (incorrect) cartoon alien after each response. No feedback was provided for the remaining 96 trials (24 abstract and 24 concrete words, plus 48 non-words), which were presented in a randomized order. Presentation of each subsequent word was prompted by the experimenter to ensure the child was on-task. To minimize fatigue, children were given the choice to take a break every 24 trials. Accuracy and reaction times were recorded; however, to ensure child attention and compliance to task instructions, the experimenter controlled stimulus presentation and did not ask the children to respond quickly, but rather as accurately as possible. Therefore, only accuracy data are analysed below. Note that this does not limit our ability to observe semantic effects, as we have shown in a previous study using the same materials and procedure [13].

(ii) Definition

Children were encouraged to provide an accurate and comprehensive definition, including as much information as they could on the meaning of each word. Each trial included the presentation of the alien in the centre of the computer screen, along with the acoustic presentation of a word. Children's responses were audio-recorded and then scored off-line but 'do not know' or definitely inaccurate responses were recorded online by the experimenter. The presentation of subsequent words was prompted by the experimenter. The 48 words were presented in four blocks of 12 items arranged in blocks corresponding to the AoA bands described previously. Words within each block were presented in random order. The task ended when the child was unable to define three words within a single block or responded to all 48 words.

Definitions were transcribed off-line and scored according to the following criteria:

(a) *Definitions' accuracy*. Definitions were scored according to the Wechsler Intelligence Scale for Children vocabulary subtest scoring criteria [33]. Scoring was performed by two independent researchers who were blind to the study hypotheses and diagnosis of the children. A third independent researcher moderated instances in which only one scorer awarded a score of 0; all other scores were averaged.

(b) *Definitions' quality ratings*. All definitions that were scored greater than 0 following the above criteria ($N = 959$) were arranged in lists of about 200 and presented to a minimum of $N = 10$ (range = 10–13) adult native English speakers, who were recruited on the crowdsourcing website Prolific (https://www.prolific.ac/). Participants were asked to rate how accurate each definition was in defining the concept. The procedure and the instructions given to raters are detailed in the electronic supplementary material. These ratings allow us to assess at a more fine-grained level the extent to which definitions of abstract words by DLD children may be of lesser quality than those by TD children.

(c) *Definitions' conceptual features*. Definitions were scored based on the 11 conceptual categories used by Barca *et al*. [34]. This classification allows us to have some initial insight on whether the conceptual features of concepts known by DLD and TD children differ. The procedure and results of this analysis are reported in the electronic supplementary material.

(d) Data analysis

DLD children were contrasted with: (i) a group of TD children matched on age (TD_{age}), to see whether DLD children had lower scores than their TD peers, especially for abstract words; (ii) a group of (younger) TD children matched on vocabulary (TD_{voc}) to further assess qualitative differences in their knowledge of concrete and abstract words. Quantitative data were analysed using mixed-effects models running in R v. 3.2.1 [35]. Lexical decision accuracy was analysed using mixed-effects logistic regression models (package 'lmerTest' [36]); definition scores were treated as ordinal and analysed using cumulative link mixed models (package 'ordinal' [37]); and average definition quality ratings were analysed using linear mixed models (package 'lmerTest' [36]). In all analyses, the baseline models included as continuous predictor the children's non-verbal reasoning scores, which significantly differed between our DLD and TD groups, and our categorical variables of interest: concreteness (abstract, concrete) and group (DLD versus TD_{age}, DLD versus TD_{voc}), as well as the two-way interaction between the two. The categorical variables were contrast-coded and the continuous predictor was centred on the mean. Log-likelihood ratio (LR) tests were used to compare fitted models. Supplementing these analyses, we performed Bayesian mixed-effects model analysis using the 'brms' package [38] for R, which fits Bayesian multilevel models using the Stan programming language. Model fitting was performed using default priors, running four chains of 10 000 iterations each. We compared models pairwise by computing the Bayes factor (BF10). A BF10 of 3 is considered sufficient evidence to favour a model over another [39,40], while a BF10 between 1/3 and 3 indicates that there is not enough evidence in the data to provide support for either model, and a BF10 $< 1/3$ indicates definite evidence against the model and in favour of the null hypothesis. Bayes factors are reported in table 3.

Qualitative data analysis of the definitions' conceptual features was carried out using correspondence analysis [41,42] running in R v. 3.2.1 (package 'CAinterprTools'; [43]), and it is reported in the electronic supplementary material.

Finally, for both lexical decision and definition tasks, case-series analyses was performed using the Revised Standardized Difference Test [44], in which the difference in performance between concrete and abstract words for each DLD child is compared with the difference in performance exhibited by the TD groups (either TD_{age} or TD_{voc}); this is reported in the electronic supplementary material.

3. Results

Below, we report the results of model selection in mixed-effects models. *P*-values from the model comparisons and corresponding Bayes factors are summarized in table 3.

(a) Lexical decision

One DLD child did not complete the task. Data from the remaining 17 children and matched controls were inspected to check whether any children showed a bias towards either answering 'word' or 'non-word'. We computed the response bias (or *criterion*, *c*), by multiplying the sum of the normalized *hit* rate (correctly identifying a word) and the normalized *false alarm* rate (incorrectly claiming that a non-word was a word) by −0.5 [45–47]. A criterion with a negative value would indicate that responses are biased towards answering 'word' (both

Table 3. Summary of *p*-values from the mixed-effects model comparisons, Bayes factors (BF10), and their interpretation. Asterisks indicate significant *p*-values or BF10 indicating either support for H0 (BF10 < 1/3) or support for H1 (BF10 > 3).

		effect	*p*-value	BF10	BF notes
lexical decision	DLD versus TD_{age}	group × concreteness	0.790	0.099*	H_0 favoured
		group	0.037*	0.85	inconclusive
		concreteness	0.622	0.172*	H_0 favoured
	DLD versus TD_{voc}	group × concreteness	0.866	0.144*	H_0 favoured
		group	0.570	0.12*	H_0 favoured
		concreteness	0.481	0.25*	H_0 favoured
definition score	DLD versus TD_{age}	group × concreteness	0.011*	17.2*	H_1 favoured
		(TD_{age}) concreteness	0.485	2.22	inconclusive
		(DLD) concreteness	0.132	5.06*	H_1 favoured
		(abstract) group	0.002*	54.6*	H_1 favoured
		(concrete) group	<0.001*	1524.9*	H_1 favoured
	DLD versus TD_{voc}	group × concreteness	0.556	0.89	inconclusive
		group	0.010*	31.2*	H_1 favoured
		concreteness	0.158	4.86*	H_1 favoured
definition quality rating	DLD versus TD_{age}	group × concreteness	0.208	0.892	inconclusive
		group	0.007*	11.68*	H_1 favoured
		concreteness	0.049*	2.13	inconclusive
	DLD versus TD_{voc}	group × concreteness	0.268	0.79	inconclusive
		group	0.010*	7.78*	H_1 favoured
		concreteness	0.051	0.78	inconclusive

*$p < 0.05$.

words and non-words are more likely to be indicated as words); a criterion of positive value would, conversely, indicate a response bias towards answering ‘non-word’ (both words and non-words are more likely to be indicated as non-words). The average criterion bias was -0.002 (s.d. $= 0.33$) for TD children, and -0.02 (s.d. $= 0.50$) for DLD children. Children who showed a criterion bias higher than 1.5 standard deviations above their group mean (indicating a strong bias towards ‘non-word’ responses) or lower than 1.5 standard deviations below their group mean (indicating a strong bias towards ‘word’ responses) were excluded from further analyses. Three children were therefore excluded from the DLD group (DLD9: $c = -0.97$; DLD12: $c = -0.74$; DLD17: $c = -0.97$); to maintain the matching between the DLD and TD groups, we also excluded the corresponding TD children.

(b) DLD versus TD_{age}

The proportion of correct responses of the two groups for concrete and abstract words is shown in figure 1*a*. We started by comparing the baseline model including the interaction between concreteness and group (see details above) against a model that included the main effects only. Including the two-way interaction did not significantly improve the fit of the model (LR for interaction model $= -542.3$; LR for main effects model $= -542.4$; $\chi^2_1 = 0.071$, $p = 0.790$).

In the main effects model, non-verbal ability (coefficient estimate $= 0.008$, s.e. $= 0.002$, $p = 0.007$) was a significant predictor of children’s performance, and it was, therefore, kept in subsequent models. We then tested whether the main effects were significant by removing them from the model, one by one. Removing the main effect of group significantly reduced the fit (LR for the model including the main effect of group $= -542.4$; LR for the model not including it $= -544.5$; $\chi^2_1 = 4.335$, $p = 0.037$), with TD_{age} children recognizing more words overall compared with DLD children (coefficient estimate $= -0.69$, s.e. $= 0.32$). Removing the main effect of concreteness did not affect the fit (LR for the model including the main effect of concreteness $= -542.4$; LR for the model not including it $= -542.5$; $\chi^2_1 = 0.243$, $p = 0.622$).

(c) DLD versus TD_{voc}

Two TD children did not complete the task owing to time constraints; therefore, they were excluded along with their matched DLD peers; this left 12 children per group. The proportion of correct responses is shown in figure 1*b*. The interaction between concreteness and group was not warranted (LR for interaction model $= -330.43$; LR for model not including it $= -330.44$; $\chi^2_1 = 0.028$, $p = 0.866$). There was no significant main effect of concreteness (LR for model including the main effect of concreteness $= -330.44$; LR for model not including it $= -330.69$; $\chi^2_1 = 0.497$, $p =$ 0.481), and no main effect of group (LR for model including the main effect of group $= -330.69$; LR for model not including it $= -330.85$; $\chi^2_1 = 0.323$, $p = 0.570$).

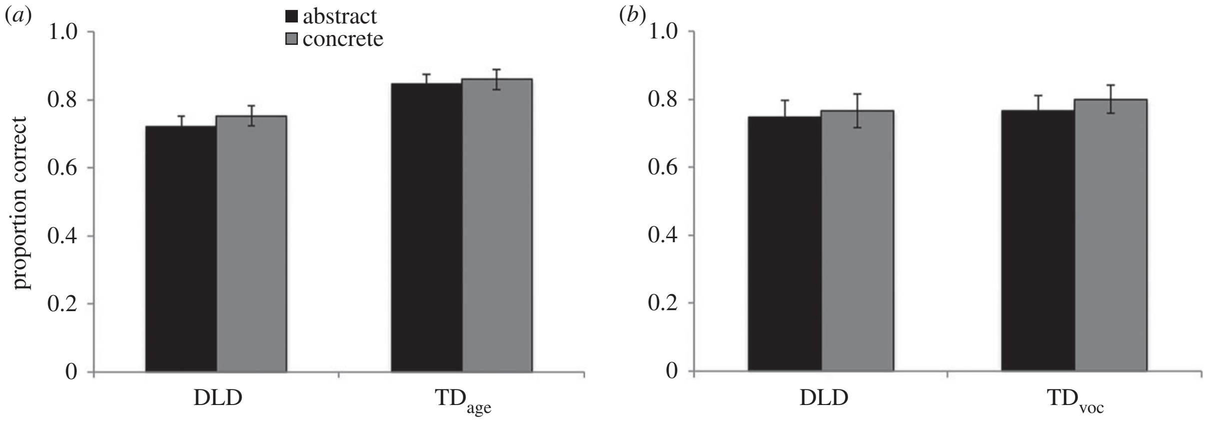

Figure 1. Proportion of correct responses to abstract and concrete words, comparing performance of DLD with TD_{age} ($N = 14$; *a*), and DLD with TD_{voc} ($N = 12$; *b*) children. Error bars indicate standard error of the mean.

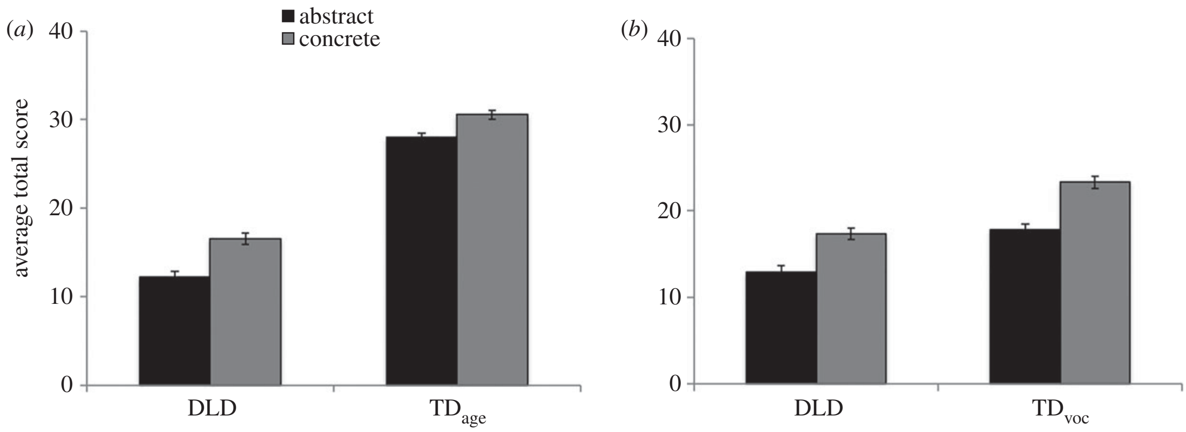

Figure 2. Average total score of definitions for abstract and concrete words, comparing performance of DLD with TD_{age} ($N = 18$; *a*), and with TD_{voc} ($N = 17$; *b*) children. Error bars indicate standard error of the mean.

(d) Definition

Only 13.4% of our TD children could provide any definition for words of AoA block 4 (words acquired at 10–11); therefore, we excluded block 4 from further analysis, reducing the total number of items to 36 words (18 abstract and 18 concrete). Overall, definitions provided by DLD children were significantly shorter (mean, $M = 7.21$ words, s.d. = 4.03) than definitions provided by TD_{age} ($M = 9.04$ words, s.d. = 7.19; $p < 0.001$) and TD_{voc} children ($M = 10.02$ words, s.d. = 8.76; $p < 0.001$), plausibly reflecting the expressive difficulties of DLD children.

(i) Definition score

DLD versus TD_{age}. Definition accuracy (raw total score) for concrete and abstract words is depicted in figure 2*a*.

Including the two-way interaction did significantly improve the fit of the model (LR for interaction model = −1191.4; LR for main effects model = −1194.6; $LR_{test} = 6.455$, $p = 0.011$). In this model, non-verbal ability (coefficient estimate = 0.15, s.e. = 0.02, $p < 0.001$) was a significant predictor of children's performance, so it was kept in subsequent models.

To interpret the significant interaction, we first looked at the main effect of concreteness separately in the two groups. We found no difference between definition scores for abstract and concrete words in both TD_{age} children (coefficient estimate = −0.45, s.e. = 0.64, $p = 0.485$) and DLD children (coefficient estimate = −1.02, s.e. = 0.67, $p = 0.132$). Looking separately at concrete and abstract words, TD_{age} children's performance was significantly better than DLD children's for both concrete words (coefficient estimate = −1.04, s.e. = 0.33, $p = 0.002$) and abstract words (coefficient estimate = −1.57, s.e. = 0.36, $p < 0.001$).

DLD versus TD_{voc}. One TD child did not complete the task and his definitions were excluded along with data from the matched DLD child. Definition accuracy (raw total score) is illustrated in figure 2*b*. The interaction between concreteness and group was not warranted (LR for interaction model = −1064.4; LR for model not including it = −1064.5; $LR_{test} = 0.346$, $p = 0.556$). The main effect of group was significant (LR for model including the main effect = −1065.5; LR for model not including it = −1068.9; $LR_{test} = 6705$, $p = 0.010$). There was no main effect of concreteness (LR for model including the main effect = −1064.5; LR for model not including it = −1065.5; $LR_{test} = 1.993$, $p = 0.158$).

(ii) Definitions' quality ratings

DLD versus TD_{age}. In the online study, we obtained ratings for 247 definitions provided by DLD children, and 439 definitions provided by their TD_{age} peers.

The concreteness × group interaction was not warranted (LR for interaction model = −1002.3; LR for main effects model = −1003.1; $\chi^2_1 = 1.583$, $p = 0.208$). The main effect of group was significant (LR for the model including the main effect of group = −1003.1; LR for the model not including it = −1006.8; $\chi^2_1 = 7.399$, $p = 0.007$), with definitions from

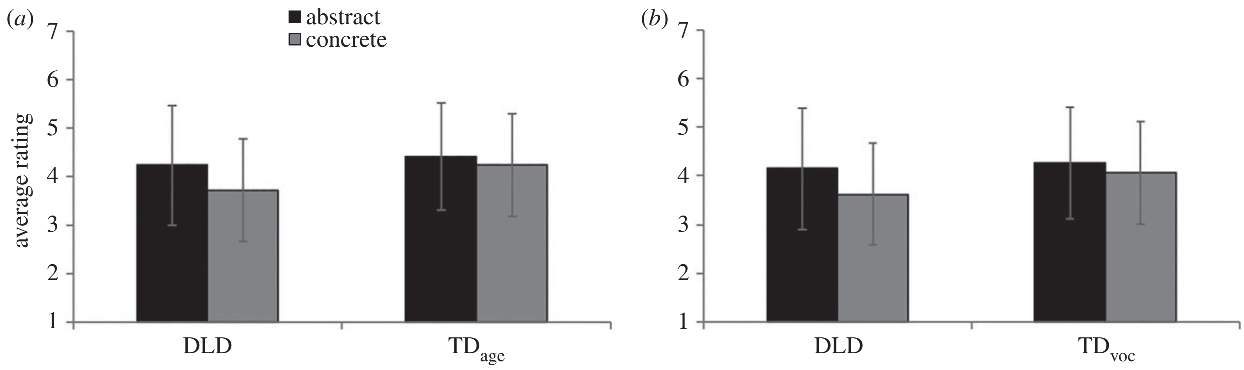

Figure 3. Average ratings (from adult native English speakers) for abstract and concrete words' definitions provided by DLD children and their matched TD_{age} peers (*a*), and by DLD children and their matched TD_{voc} peers (*b*). Error bars indicate standard deviations.

TD_{age} children rated as more accurate overall than those provided by their DLD peers. The main effect of concreteness was also significant (LR for the model including the main effect of concreteness = −1003.1; LR for the model not including it = −1001.1; $\chi^2_1 = 3.963$, $p = 0.047$). Crucially, definitions of abstract words were rated as more accurate than definitions of concrete words, for both DLD and TD_{age} children (figure 3*a*).

DLD versus TD_{voc}. We analysed ratings for 244 definitions provided by DLD children, and 314 definitions provided by their TD_{voc} peers.

The concreteness × group interaction was not warranted (LR for interaction model = −810.4; log-likelihood ratio for main effects model = −811.0; $\chi^2_1 = 1.123$, $p = 0.268$). The main effect of group was significant (log-likelihood ratio for the model including the main effect of group = −810.4; LR for the model not including it = −814.3; $\chi^2_1 = 6.667$, $p = 0.010$), with definitions of TD_{voc} children rated as more accurate than those provided by their DLD peers. The main effect of concreteness was marginally significant (LR for the model including the main effect of concreteness = −810.4; LR for the model not including it = −812.9; $\chi^2_1 = 3.817$, $p = 0.051$), with definitions of abstract words rated as more accurate than definitions of concrete words (figure 3*b*).

4. Discussion

This study aimed to assess whether linguistic development has a greater role—a primacy—in the learning of abstract compared with concrete concepts as predicted by theories such as dual coding [3] and context availability [4]. We tested knowledge of abstract and concrete words in children with DLD and age-matched as well as (younger) vocabulary-matched peers, using both a lexical decision and a definition task.

In the lexical decision task, we found that children with DLD recognized significantly fewer words overall compared with their age-matched TD peers; however, this was a small effect and not confirmed by the Bayes factor analysis. What is of most interest here, however, is that while DLD children's performance was impaired with all words, they did not show a disproportionate impairment with abstract words compared with concrete, as confirmed by the lack of a concreteness by group interaction, supported by a Bayes factor in favour of the null hypothesis. Interestingly, when looking at the comparison between DLD children and their vocabulary-matched peers, we found no significant differences at all. The lexical decision task, however, only gives us an indication of how many words children could recognize, and it cannot tell us anything about children's appreciation of word meaning.

In the definition task, when looking at definition accuracy, we do find a significant interaction between concreteness and group. What the results of the definition task suggest is that TD_{age} children give more accurate definitions compared with DLD children for both abstract and concrete words, and they define abstract words with similar accuracy to concrete words (although Bayes factor analysis suggests there is no enough evidence to accept the null hypothesis), while children with DLD show a larger difference between accuracy for abstract and concrete definitions. When compared with younger TD_{voc} children, children with DLD are worse at defining all words, not only abstract words. The additional analyses reported in the electronic supplementary material, which contrast the difference in performance between abstract and concrete words for each individual DLD child against the average difference shown by TD_{age} and TD_{voc} children, confirm that any difference in definition scores for abstract and concrete words is equivalent to that exhibited by the TD groups.

When we look at the quality ratings that adults provided of how accurate each definition is in defining the concept, we find again strong support for a difference between the quality of definitions provided by TD children (both TD_{age} and TD_{voc}) and children with DLD, but the lack of an interaction between group and concreteness suggests that the difference, if any, between DLD children's definitions of abstract and concrete concepts is not significantly larger than the difference, if any, exhibited by TD children. Interestingly, although the Bayes factor suggests the evidence to argue for or against a main effect of concreteness is inconclusive, the marginally significant *p*-values suggest that adults rated definitions of abstract words (overall) as slightly more accurate than definitions of concrete words. This is an unexpected but interesting result which might be linked to task expectations. Adult raters were recruited over the Internet and they knew that the definitions were provided by children. It may be that, in general, they were less strict for the abstract words as these are typically considered to be harder for children.

To summarize, DLD children show impaired performance at recognizing both abstract and concrete words compared with their age-matched peers, but they can correctly recognize an equivalent number of words (both abstract and concrete) compared with their vocabulary-matched peers (as showed by the lexical decision task). However, they cannot provide the same level of quality of definitions. It is worth noting here that TD_{voc} children were matched to our DLD children on receptive vocabulary scores, but the definitions task requires expressive language skills, which are impaired in children with DLD (as also supported by the fact that their definitions were shorter than those of TD children).

Taken together, our lexical decision results as well as the results from analysis of definition accuracy, definition quality ratings and the comparisons of individual DLD children with the TD groups, do not provide clear support for linguistic primacy in the learning of abstract words and concepts. When language development is impaired, as is the case for DLD children, knowledge of both abstract and concrete words is impaired. When expressive vocabulary is not required, children with DLD perform like younger TD children with equivalent receptive vocabulary. This suggests that the same factors might support the learning of new words in young children and children with DLD.

A number of theoretical accounts assume that embodied information contributes to the semantic representation of words. For example, Kousta *et al*. [11] suggested that while words referring to concrete objects and actions would be learnt by associating sensory-motor experience with the word, abstract words would be learnt by associating emotional states with the word. Ponari *et al*. [13] showed that TD children up to the age of 8–9 (about the age range of our TD_{voc} children) have better knowledge of emotionally valenced abstract words. They suggested that emotion might be particularly important for the acquisition of abstract words early in childhood, when the vocabulary is mainly acquired through social interactions, providing a bootstrapping mechanism. Emotional valence could support children in discovering that some words—those that trigger emotional reactions—refer to internal states, rather than to objects and actions in the environment, thus providing the building blocks for establishing the general category of abstract concepts. Later on, after the age of 9, the effect of valence declines. They suggested that as vocabulary and linguistic competence increases, children make greater use of linguistic information (e.g. from the text), and are more able to make use of correlational patterns in discourse in order to extrapolate abstract meaning from the linguistic context [13]. Children with DLD have reduced vocabulary and deficits in syntactic competence, and it has been shown that they are not as attuned as TD peers to statistical co-occurrences in language input [48]. However, they do not have sensory-motor, or emotional/social impairments. Thus, they can benefit from the same embodied mechanisms for learning both concrete and abstract words as their TD peers.

The qualitative analysis of the content of the definitions provided by DLD as well as TD children (both TD_{age} and TD_{voc}) reported in the electronic supplementary material supports the idea that sensory-motor associations are crucial for concrete words while affective associations are crucial for abstract words [11]. Here, we found that definitions of concrete concepts include more perceptual features of the referent, their spatial location or function, as well as superordinate levels of the taxonomy, while abstract concepts' definitions include more situational and emotional features. These different features provide a clear distinction between abstract and concrete categories, at least in TD children. According to the same analysis, however, definitions of abstract and concrete words in DLD children are less clearly distinct, and it seems less clear whether children with DLD make use of embodied emotional and situational features when defining abstract concepts. In summary, while children with DLD do not seem to be more impaired for abstract versus concrete words compared with their TD peers in terms of how accurate their definitions are, DLD children might not use embodied (sensory-motor, emotional and situational) information to the same extent as their TD peers. However, we can only speculate on the basis of the current data as the differences might just reflect expressive deficits of the DLD children.

In conclusion, the study presented the first investigation of abstract word knowledge by children with DLD. Our results confirm the role of linguistic information in the representation of concepts across domains of knowledge. Children with DLD show poorer vocabulary when compared with age-matched TD children. We do not support, however, a special role for linguistic information in the learning of abstract concepts. It is for future studies to further investigate to what extent children with DLD can take advantage of sensory-motor and emotional information in learning the semantics of both concrete and abstract words.

Data accessibility. Data I have been deposited at: http://data.kent.ac.uk/21/. The doi link for the dataset is http://doi.org/10.22024/UniKent/01.01/21.

Competing interests. We declare we have no competing interests.

Authors' contributions. G.V. and C.F.N. designed and coordinated the study; M.P. participated in the design of the study, collected the data, carried out the statistical analyses; A.R. and A.L. contributed to interpretation; all authors contributed to draft the manuscript and gave final approval for publication.

Funding. This research was supported by a grant from the Nuffield Foundation (EDU/40477) to G.V. and C.F.N.

Disclaimer. The views expressed are those of the authors and not necessarily those of the Foundation.

References

1. Medina TN, Snedeker J, Trueswell JC, Gleitman LR. 2011 How words can and cannot be learned by observation. *Proc. Natl Acad. Sci. USA* **108**, 9014–9019. (doi:10.1073/pnas.1105040108)
2. Binder JR, Desai RH. 2011 The neurobiology of semantic memory. *Trends Cogn. Sci.* **15**, 527–536. (doi:10.1016/j.tics.2011.10.001)
3. Paivio A. 2007 Mind and its evolution: a dual coding theoretical approach. *Mahwah* **63**, 250–266. (doi:10.4324/9781315785233)
4. Schwanenflugel PJ. 1991 Why are abstract concepts hard to understand? In *The psychology of word meanings* (ed. PJ Schwanenflugel), pp. 223–250. Hillsdale, NJ: Lawrence Erlbaum Associates.
5. Hoffman P, Binney RJ, Lambon Ralph MA. 2015 Differing contributions of inferior prefrontal and anterior temporal cortex to concrete and abstract conceptual knowledge. *Cortex* **63**, 250–266. (doi:10.1016/j.cortex.2014.09.001)
6. Recchia G, Jones MN. 2012 The semantic richness of abstract concepts. *Front. Hum. Neurosci.* **6**, 315. (doi:10.3389/fnhum.2012.00315)
7. Meteyard L, Cuadrado SR, Bahrami B, Vigliocco G. 2012 Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex* **48**, 788–804 (doi:10.1016/j.cortex.2010.11.002)
8. Borghi AM, Binkofski F, Castelfranchi C, Cimatti F, Scorolli C, Tummolini L. 2017 The challenge of abstract concepts. *Psychol. Bull.* **143**, 263–292. (doi:10.1037/bul0000089)

9. Moseley R, Carota F, Hauk O, Mohr B, Pulvermüller F. 2012 A role for the motor system in binding abstract emotional meaning. *Cereb. Cortex* **22**, 1634–1647. (doi:10.1093/cercor/bhr238)
10. Barsalou LW, Wiemer-Hastings K. 2005 Situating abstract concepts. In *Grounding cognition: the role of perception and action in memory, language and thought* (eds D Pecher, R Zwaan), pp. 129–163. New York, NY: Cambridge University Press.
11. Kousta S-T, Vigliocco G, Vinson DP, Andrews M, Del Campo E. 2011 The representation of abstract words: why emotion matters. *J. Exp. Psychol. Gen.* **140**, 14–34. (doi:10.1037/a0021446)
12. Vigliocco G, Kousta S-T, Della Rosa PA, Vinson DP, Tettamanti M, Devlin JT, Cappa SF. 2014 The neural representation of abstract words: the role of emotion. *Cereb. Cortex* **24**, 1767–1777. (doi:10.1093/cercor/bht025)
13. Ponari M, Norbury C, Vigliocco G. 2017 Acquisition of abstract concepts is influenced by emotional valence. *Dev. Sci.* **21**, e12549. (doi:10.1111/desc.12549)
14. Dove G. 2009 Beyond perceptual symbols: a call for representational pluralism. *Cognition* **110**, 412–431. (doi:10.1016/j.cognition.2008.11.016)
15. Andrews M, Vigliocco G, Vinson D. 2008 Integrating experiential and distributional data to learn semantic representations. *Psychol. Rev.* **116**, 463–498. (doi:10.1037/a0016261)
16. Norbury CF, Gooch D, Wray C, Baird G, Charman T, Simonoff E, Vamvakas G, Pickles A. 2016 The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *J. Child Psychol. Psychiatry* **57**, 1247–1257. (doi:10.1111/jcpp.12573)
17. Rice ML. 2013 Language growth and genetics of specific language impairment. *Int. J. Speech Lang. Pathol.* **15**, 223–233. (doi:10.3109/17549507.2013.783113)
18. McGregor KK, Oleson J, Bahnsen A, Duff D. 2013 Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *Int. J. Lang. Commun. Disord.* **48**, 307–319. (doi:10.1111/1460-6984.12008)
19. Henry LA, Botting N. 2017 Working memory and developmental language impairments. *Child Lang. Teach. Ther.* **33**, 19–32. (doi:10.1177/0265659016655378)
20. Romberg AR, Saffran JR. 2010 Statistical learning and language acquisition. *Wiley Interdiscip. Rev. Cogn. Sci.* **1**, 906–914. (doi:10.1002/wcs.78)
21. Ebert KD, Kohnert K. 2011 Sustained attention in children with primary language impairment: a meta-analysis. *J. Speech Lang. Hear. Res.* **54**, 1372–1384. (doi:10.1044/1092-4388(2011/10-0231))
22. Dunn LM, Dunn LM, Whetton C, Burley J. 1997 *British picture vocabulary scale*, 2nd edn. Windsor, UK: NFER-Nelson.
23. Wechsler D. 2011 *Wechsler abbreviated scale of intelligence–second edition (WASI-II)*. San Antonio, TX: NCS Pearson.
24. Semel E, Wiig EH, Secord WA. 2006 *Clinical evaluation of language fundamentals*, 4th edn. San Antonio, TX: PsychCorp.
25. Kuperman V, Stadthagen-Gonzalez H, Brysbaert M. 2012 Age-of-acquisition ratings for 30 000 English words. *Behav. Res. Methods* **44**, 978–990. (doi:10.3758/s13428-012-0210-4)
26. Brysbaert M, Warriner AB, Kuperman V. 2014 Concreteness ratings for 40 thousand generally known English word lemmas. *Behav. Res. Methods* **46**, 904–911. (doi:10.3758/s13428-013-0403-5)
27. Warriner AB, Kuperman V, Brysbaert M. 2013 Norms of valence, arousal, and dominance for 13 915 English lemmas. *Behav. Res. Methods* **45**, 1191–1207. (doi:10.3758/s13428-012-0314-x)
28. Balota DA *et al.* 2007 The English lexicon project. *Behav. Res. Methods* **39**, 445–459. (doi:10.3758/BF03193014)
29. van Heuven WJB, Mandera P, Keuleers E, Brysbaert M. 2014 SUBTLEX-UK: a new and improved word frequency database for British English. *Q. J. Exp. Psychol.* **67**, 1176–1190. (doi:10.1080/17470218.2013.850521)
30. Coltheart M. 1981 The MRC psycholinguistic database. *Q. J. Exp. Psychol. A* **33**, 497–505.
31. Audacity Team X. *Audacity*®. Pittsburgh, PA: Carnegie Mellon University.
32. Psychology Software Tools. 2012 *E-Prime 2.0*. See https://www.pstnet.com.
33. Wechsler D. 2014 *Wechsler intelligence scale for children*, 5th edn. Bloomington, MN: Pearson.
34. Barca L, Mazzuca C, Borghi AM. 2017 Pacifier overuse and conceptual relations of abstract and emotional concepts. *Front. Psychol.* **8**, 2014. (doi:10.3389/fpsyg.2017.02014)
35. R Core Team. 2017 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. See https://www.R-project.org/.
36. Kuznetsova A, Brockhoff P, Christensen R. 2017 lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* **82**, 1–26. (doi:10.18637/jss.v082.i13)
37. Christensen R. 2018 *ordinal – Regression models for ordinal data*. R package version 2018.4-19. See http://www.cran.r-project.org/package=ordinal/.
38. Bürkner P-C. 2017 brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28. (doi:10.18637/jss.v080.i01)
39. Jeffreys H. 1998 *Theory of probability*. Oxford, UK: Clarendon Press.
40. Kass RE, Raftery AE. 1995 Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795. (doi:10.1080/01621459.1995.10476572)
41. Greenacre MJ, Blasius J. 1994 *Correspondence analysis in the social sciences: recent developments and applications*. London, UK: Academic Press.
42. Greenacre MJ. 2016 *Correspondence analysis in practice*. Boca Raton, FL: CRC Press.
43. Alberti G. 2015 CAinterprTools: an R package to help interpreting correspondence analysis' results. *SoftwareX* **1–2**, 26–31. (doi:10.1016/j.softx.2015.07.001)
44. Crawford JR, Garthwaite PH. 2005 Evaluation of criteria for classical dissociations in single-case studies by Monte Carlo simulation. *Neuropsychology* **19**, 664–678. (doi:10.1037/0894-4105.19.5.664)
45. Fox JR. 2004 A signal detection analysis of audio/video redundancy effects in television news video. *Communic. Res.* **31**, 524–536. (doi:10.1177/0093650204267931)
46. Macmillan NA, Creelman CD. 2004 *Detection theory: a user's guide*. Cambridge, UK: Cambridge University Press.
47. Shapiro MA. 1994 Signal detection measures of recognition memory. In *Measuring psychological responses to media messages* (ed. A Lang), pp. 133–148. Hillsdale, NJ: Lawrence Erlbaum Associates
48. Evans JL, Saffran JR, Robe-Torres K. 2009 Statistical learning in children with specific language impairment. *J. Speech Lang. Hear. Res.* **52**, 321–335. (doi:10.1044/1092-4388(2009/07-0189))