Individual differences in autistic children’s homograph reading: Evidence from Hebrew

Jon Brock
ARC Centre of Excellence in Cognition and its Disorders, Australia; Department of Cognitive Science, Macquarie University, Australia; Department of Psychology, Macquarie University, Australia

Nufar Sukenik
Language and Brain Lab, School of Education and Sagol School of Neuroscience, Tel Aviv University, Israel

Naama Friedmann
ARC Centre of Excellence in Cognition and its Disorders, Australia; Language and Brain Lab, School of Education and Sagol School of Neuroscience, Tel Aviv University, Israel

Abstract

Background and aims: On average, autistic individuals make more errors than control participants when reading aloud sentences containing heterophonic homographs—written words with multiple meanings and pronunciations. This finding is widely interpreted within the framework of “weak central coherence” as evidence for impaired sentence-level comprehension resulting in a failure to disambiguate the homograph meaning. However, consistent findings at the group level belie considerable individual variation. Our aim here was to determine whether that variation was reliable and whether it could be predicted.

Methods: We developed a Hebrew version of the homograph-reading test, containing many more items than is possible in English. The test was administered to 18 native-Hebrew speaking autistic children and adolescents, along with a battery of reading and language assessments.

Results: Participants with autism showed wide individual variation in performance on the homograph-reading task. Using a mixed random effects logistic regression analysis, we showed that measures of autism severity, single word reading, and single word comprehension all left reliable individual variation unaccounted for and none accounted for variation beyond that associated with the child’s age. Instead, homograph reading was best predicted by performance on a picture naming task, which accounted for unique variation beyond age and each of the other predictors.

Conclusions: Poor performance of autistic individuals on the English version of the homograph-reading task has until now been characterized as evidence for a comprehension deficit in autism. However, the results of the current study lead us to propose a new working hypothesis—that difficulties affecting some autistic individuals reflect impairment in the use of semantics to guide the selection of the appropriate phonological form during speech production. This hypothesis is consistent with the strong association between homograph reading and picture naming. It may also help explain the inconsistent pattern of results across studies using different measures of linguistic “central coherence.”

Implications: The results of this preliminary study should be replicated before firm conclusions are drawn. Nonetheless, the study serves to emphasize the importance of considering within-group as well as between-group variations in studies of autism. It also provides a worked example showing how mixed random effect analyses can be used to explore individual differences, distinguishing between genuine variation and psychometric noise.

Keywords

Autism spectrum disorders, expressive language, reading, semantics

Corresponding author:
Naama Friedmann, Language and Brain Lab, School of Education and Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel.
Email: naamafr@post.tau.ac.il

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Introduction

“Autism spectrum disorder” (ASD) is a diagnostic label given to individuals who exhibit social and communication difficulties co-occurring with repetitive behaviors or restricted interests (American Psychiatric Association, 2000, 2013). Current criteria allow for considerable heterogeneity of behavioral symptoms, such that two individuals with the same diagnosis may have little if anything in common beyond the autism label itself. This heterogeneity is widely acknowledged as a challenge for autism research (Georgiades, Szatmari, & Boyle, 2013). Indeed, many researchers now talk about “the autisms” plural (Geschwind & Levitt, 2007) and some even question whether “autism” is a useful scientific concept at all (Waterhouse, 2012).

However, the default paradigm for investigating autistic cognition remains the comparison of groups of participants with and without an autism diagnosis, treating individual variation as essentially meaningless noise (Brock, 2011).

A pertinent illustration of these issues comes from one of the most influential and well-replicated findings in autism research. In 1983, Frith and Snowling asked six autistic children to read aloud sentences containing heterophonic homographs—words such as “row” and “tear” that have multiple meanings and pronunciations associated with the same orthographic form. The word “tear”, for example, is pronounced differently in the phrases “In her eye was a tear” versus “In her dress was a tear.” The autistic children performed significantly worse than control children, tending to give the more common pronunciation of the homograph regardless of semantic cues earlier in the sentence.

Frith and Snowling concluded, therefore, that autism is characterized by “a failure to utilize semantic context” (p. 339) during language comprehension. Subsequently, this finding became a key piece of evidence for Frith’s “weak central coherence” account, according to which many behavioral symptoms associated with autism reflect a tendency to process information out of context (Frith, 1989).

Frith and Snowling’s original finding has since been replicated on at least four occasions (Burnette et al., 2005; Happé, 1997; Jolliffe & Baron-Cohen, 1999; López & Leekam, 2003; but see Snowling & Frith, 1986).1 These studies have confirmed that, as a group, autistic adults and children make more errors on the homograph-reading task than control participants, even when basic reading skills, vocabulary knowledge, and IQ are controlled for.

However, the significant group differences in performance belie considerable individual variation. The critical condition of the test occurs when the preceding context is consistent with the subordinate (less common) meaning of the homograph and so a correct response can only be ascertained by taking account of the context. Happé (1997), Jolliffe and Baron-Cohen (1999), and Lopez and Leekam (2003) all reported that their autism group performed significantly below their control group in this condition. Yet a closer inspection of the data reveals that, of a total 65 autistic participants tested, 26 performed without error in this condition and a further 19 made only one error (Brock & Caruana, 2014). The interesting question, therefore, is not why in general people with autism perform poorly on the homograph-reading task, but why some individuals appear to struggle whereas others do not. In other words, there is a need to look at individual differences within the autism population as well as group differences between autistic and nonautistic participants.

Unfortunately, the investigation of individual differences in homograph reading is severely limited by the fact there are only four English homographs appropriate for the task.2 Other potential homographs are ruled out because they have the same pronunciation for both meanings (e.g., “bank”); because one or both of the meanings is likely to be unfamiliar to some participants (e.g., “sewer”); or because the two meanings come from different syntactic classes, thereby providing nonsemantic clues to their meaning (e.g., “wind,” “present”). With only a small number of trials per condition, it becomes unclear to what extent variation in performance across participants reflects genuine individual differences in ability as opposed to mere psychometric noise.

Given the limitations of the English version of the homograph-reading task, in the current study we devised a similar test in Hebrew, a language characterized by a much greater degree of orthographic ambiguity. Hebrew has 22 letters, 4 of which can be mapped onto two different consonantal sounds, and 5 can serve either as a vowel (or several vowels) or as a consonant. Some vowels are rarely represented in writing, and even when a vowel is represented, the corresponding letter is usually ambiguous between several vowels and consonants. Finally, as in English, stress in Hebrew is not represented in the orthography, and stress position is lexically specified. As a result, there are a large number of candidate heterophonic homographs in Hebrew, many of them frequent and known to children in both meanings. We were therefore able to include a much larger number of trials than has been possible in previous studies with English-speaking participants.

As in these previous studies, we asked young people with autism to read aloud heterophonic homographs that were incorporated in sentences that biased their meaning toward only one reading but did not provide a syntactic cue to its appropriate reading. However, instead of comparing autistic and non-autistic children,
our aim was to investigate the variation in performance within the autism group. To this end, we subjected participants’ responses on the homograph task to logistic regression with mixed random effects, treating each trial as a binary data point (correct or incorrect), and modeling the likelihood of a correct response on each trial based on characteristics of both the participant and the item in question (Jaeger, 2008). An advantage of this approach over more conventional regression analyses (which average across items for each participant) is that it provides an indication of the reliability of individual variation. That is, we were able to determine the extent to which there is genuine variation in performance across participants as opposed to mere psychometric noise.

The next step was to try and explain the participant-related variance. If poor homograph reading is characteristic of autism (as suggested by the weak central coherence account), then we might expect individuals with more severe autism symptoms to have greatest difficulty. However, poor performance might also be attributed to more general reading or language difficulties (cf. Norbury & Nation, 2011). For example, children with poor decoding skills would be expected to make more errors on any reading task; children with limited vocabularies might be less familiar with the different homograph meanings and perhaps also the meanings of critical contextual words; children with lexical retrieval difficulties might struggle to access the appropriate word meanings and output the correct words accordingly. Thus, in addition to performance on the homograph task itself, we also measured children’s reading of individual (nonhomographic) words, their comprehension of individual written words, and their picture naming—a measure of lexical retrieval abilities. By adding these measures to the logistic regression model, we could determine their relative importance in accounting for variation in homograph reading.

**Method**

**Participants**

Participants in this preliminary study were 18 autistic children and adolescents (16 boys) aged 9.0 to 18.0 years, who were taking part in a larger study of language skills in ASD at Tel Aviv University. Nine of the participants were in 9th to 11th grade, and nine were in 3rd to 5th grade. All participants spoke Hebrew as their native language. It was also the first language they learned to read and the language in which they studied at school. Three further children were excluded due to inadequate reading skills, defined as failure to complete the TILTAN test of single word reading (see below for details) or a score below 50% on that test. Details of the final 18 participants are provided in Table 1.

All the participants were diagnosed with autistic disorder by a child psychiatrist prior to the study according to the DSM-IV criteria (American Psychiatric Association, 2000), and were recommended for an ASD-specific class. Seventeen of the participants were enrolled in autism-specific classes, and the remaining child received 1:1 support in a mainstream class. Thirteen were described as “high functioning.”

**Background measures**

The participants were tested, as part of a larger study, on a battery of language and reading tests. This contained a number of assessments that were potentially relevant to homograph reading and were therefore included as predictors of individual variation in performance.

**CARS.** To provide an index of autism severity, the Childhood Autism Rating Scale (CARS; Schopler, Reicher, & Rochen Renner, 1988) was completed by each child’s teacher.

**Single word reading.** As a measure of basic decoding skills at the word level, participants were administered the word reading screening part of the TILTAN reading test (Friedmann & Gvion, 2003). The TILTAN has a reliability coefficient of 0.968 (Kuder & Richardson, 1937) based on a sample of 1022 children with and without reading difficulties. It was developed to detect all known types of decoding difficulties, by using words that are sensitive to the various types of errors (irregular words, migratable words, morphologically complex words, words in which a letter error creates other existing words, etc.). Participants read aloud 136 single words. For each participant, we calculated the proportion of words read correctly and the type of errors made.

**Word association.** Knowledge of word meaning was assessed using the MA KASHUR word association test (Biran & Friedmann, 2007). The test includes 35

| Table 1. Participant characteristics. |
|---------------------------------------|
| Min | Max | Mean | SD |
|-----|-----|------|----|
| Age (years) | 9.0 | 18.0 | 13.3 | 3.1 |
| CARS (total score) | 24.0 | 43.5 | 32.7 | 5.8 |
| TILTAN single word reading (% correct) | 53 | 98 | 82 | 15 |
| MA KASHUR word association (% correct) | 57 | 100 | 88 | 10 |
| SHEMESH picture naming (% correct) | 73 | 100 | 91 | 6 |
triads of written words—a target word in the top half of a display (e.g., “soup”), with two additional words presented underneath that are semantically related to each other (e.g., “spoon” and “fork”). The participants were required to decide which of the two words was more closely related to the target word. Twenty-five of the target words were imageable nouns. The remaining 10 items were abstract nouns and adjectives.

**Picture naming.** Lexical retrieval abilities were assessed using the SHEMESH picture naming test (Biran & Friedmann, 2005). Participants were asked to name 100 color pictures of objects of various semantic categories (animals, tools, fruit, vegetables, vehicles, musical instruments, utensils, appliances, apparel, jewelry, body parts) and were scored on the proportion of immediate correct responses. The target words were one to four syllable nouns, with ultimate and penultimate stress and a range of initial phonemes; they included both masculine and feminine nouns, with regular and irregular gender morphology. Mean frequency of these nouns is 81 occurrences in a million (SD = 251; range 1–2006) based on a corpus of 165 million words from the Israblog website (Linzen, 2009). Reliability is 0.836 based on a sample of 1022 participants with and without reading difficulties. Performance of typically developing children is close to ceiling (mean = 94.1%; SD = 2.3% for 35 12- to 14-year olds; Biran & Friedmann, 2005).

**The homograph-reading task**

Sixteen heterophonic homographs were each incorporated into two sentences such that only one interpretation was semantically acceptable. Importantly, the sentences contained no syntactic cues to the appropriate pronunciation of the homograph. As in Frith and Snowling’s original study, the homograph appeared toward the end of the sentence after the disambiguating context. The example below shows the English gloss of two sentences presented with the homograph קופה, which could be read as either /kupa/, meaning cashier, or /kofa/, meaning (female) monkey:

(1) In the mall, mom bought a dress and she had to pay, so she looked for a cashier.
(2) In the zoo, the guide showed everyone a giraffe and a zebra and then she also found a sweet monkey.

Test sentences were divided into two blocks administered on different days, with each homograph appearing once in each block to minimize potential interference effects (cf. Brock & Bzishvili, 2013). The participants were asked to read the sentences aloud with no time limit. The experimenter gave general encouragement but no feedback with regard to success on individual test items.

Responses were recorded and coded during the session and were recoded independently offline with any disagreements resolved by discussion. Here, we report analyses for the initial pronunciation of the homograph (cf. Frith & Snowling, 1983), although outcomes in terms of significance were identical if we allowed for self-correction. Unlike previous studies, we made no distinction between dominant and subordinate pronunciations of the homograph as dominance varies depending on the individual participant’s language experience. Importantly, given that the participants are reading the homograph in two sentences, each biasing toward a different meaning and pronunciation, if they do not use the sentence meaning to select the pronunciation, they would be expected to make at least one error in every pair of homographs, regardless of which meaning is dominant for them.

In parallel with the testing of autistic children for this study, the homograph-reading task was also administered to 14 typically developing children in fourth grade (mean age = 10.3, SD = 0.3). For three of the homographs it became clear that children were unfamiliar with one of the meanings. The six sentences involving these three homographs were, therefore, excluded from analysis of the autistic participants’ data. The remaining 13 homograph pairs, which we included in the test and analyses, are provided in the Appendix (Table 4). The typically developing fourth graders read the homographs in these 26 sentences almost perfectly. One child made five homograph reading errors, but the other 13 children made at most one error each.

**Analysis**

Data from the homograph task were subjected to logistic regression analyses with mixed random effects (Baayen, Davidson, & Bates, 2008) using the lme4 package (Bates, 2015) in R version 3.3.2. Rather than taking each participant’s total correct trials as the dependent variable, this approach considers each trial as an individual data point and attempts to model the likelihood of a correct response on any given trial, treating both subject and item as random effects. Analyses involved nested-model comparisons based on likelihood ratio tests. Predictors (e.g., age, word association scores) were added to the model as fixed factors and we determined whether the more complex model provided a significantly better fit to the data (α = .05). Analysis code and output are provided in the R Markdown (online Supplementary Material B).
Results

The results of the homograph-reading test indicated that some individuals with autism showed considerable difficulty in reading the heterophonic homographs, whereas others showed near-ceiling levels of performance. As shown in the R Markdown (online Supplementary Material), the majority of errors (67%) were classified as homograph errors, whereby a contextually inappropriate homograph was provided. A further 16% were classified as surface errors—non-words that formed phonologically plausible alternative readings of the homographs. Equivalent errors in English would be reading the word walk with a pronounced “l,” or reading the word wear rhyming with “dear.” The remaining 17% of errors were orthographically illegal readings of the homographs.

Initial analyses (see R Markdown) showed that, unsurprisingly, older children performed better than younger children. Thus, the analyses we report here considered whether there was a significant association between homograph reading and each predictor, beyond the shared influence of age on both measures.

To evaluate the strength of association between homograph reading and each of the predictors, we built a simple model with age and a second predictor (e.g., word reading accuracy) as fixed factors and subject and item as random factors. This is illustrated in Model 1 below. The tilde (~) denotes “is modeled as a function of . . . .” Random effects are distinguished from fixed effects by the prefix “1|” indicating a random intercept. Figure 1 shows homograph reading accuracy plotted against each of the predictors and the model fit, as well as the predicted values for each participant based on the model.

Model 1: Correct ~ Word Reading + Age + (1|Subject) + (1|Item)
Model 2: Correct ~ Age + (1|Subject) + (1|Item)
Model 3: Correct ~ Word Reading + Age + (1|Item)

We then determined whether removing the predictor (Model 2) significantly reduced model fit. As shown in Table 2 (left column), this was only true of picture naming. In other words, picture naming was the

| Predictor               | Subject   | \(\chi^2\) (1) | \(p\)  | \(\chi^2\) (1) | \(p\)  |
|-------------------------|-----------|---------------|-------|---------------|-------|
| CARS                    |           | 2.966         | .085  | 4.334         | .037  |
| TILTAN single word reading |         | 1.953         | .162  | 7.471         | .006  |
| MA KASHUR word association |         | 2.876         | .090  | 5.298         | .021  |
| SHEMESH picture naming  |           | 6.860         | .009  | 2.170         | .141  |

Predictor column shows the effect of removing the predictor (fixed factor) from the full model, which includes that single predictor, age, and subjects and items as random factors. The \(p\)-value here indicates the statistical significance of the predictor. The Subject column shows the effect of removing the by-subjects random factor from the full model. Here, the \(p\)-value indicates whether or not there is reliable by-subjects variation unexplained by the full model.

Figure 1. Performance of individual participants on the homograph-reading task plotted against each of the predictors. Black data points show the proportion correct for each participant. Orange logistic functions represent the contribution of the predictor as determined by the corresponding fixed effect in the mixed random effects model (Model 1 in main text). Orange data points show the age-adjusted predicted value for each participant based on the model (conditional modes). Error bars are prediction intervals.
only predictor accounting for significant variation beyond age.

For each predictor, we also examined the effect of removing the by subjects random factor from the model (i.e., Model 3 vs. Model 1). For CARS score, word reading score, and word association score, this significantly reduced model fit (see right column of Table 2), indicating that there was reliable by-subjects variation unaccounted for. However, for picture naming, model fit was not significantly reduced, indicating that there was little reliable individual variation that was not accounted for by picture naming.

Given the strong association with picture naming, we performed additional model comparisons (Table 3). These showed that picture naming accounted for unique variation beyond that accounted for by each of the other predictors (Model 4 vs. Model 5). In contrast, none of the other predictors accounted for variation beyond picture naming (Model 4 vs. Model 6).

Model 4: Correct ~ Word Reading + Picture Naming + (1|Subject) + (1|Item)
Model 5: Correct ~ Word Reading + (1|Subject) + (1|Item)
Model 6: Correct ~ Picture Naming + (1|Subject) + (1|Item)

Further analyses (see R Markdown in online Supplementary Material) showed that the effect of Picture Naming was not driven by outliers with excessive leverage and that similar results were obtained when age was treated as a group variable (younger vs. older children) rather than a continuous variable.

Finally, we performed more conventional regression analyses of the homograph reading performance taking the logit-transformed proportion correct for each participant as the dependent variable. Consistent with the logistic regression analyses, only picture naming accounted for significant variation beyond age.

**Table 3. Results of mixed random effects logistic regression analyses.**

| Picture naming Covariate | \( \chi^2 \) (1) | \( p \) |
|----------------------------|-----------------|---------|
| Age                        | 6.860           | .009    |
| CARS                       | 9.120           | .003    |
| TILTAN single word reading | 6.977           | .008    |
| MA KASHUR word association | 10.105          | .001    |

For each analysis, the full model includes picture naming and one other predictor (covariate) as fixed factors. The picture naming column shows the effect of removing picture naming from the full model. The \( p \)-values indicate that picture naming accounts for significant variation beyond each covariate. The covariate column shows the effect of removing the covariate from the full model. Here, the \( p \)-values indicate that none of the covariates account for significant variation beyond picture naming.

Discussion

Autism, as currently defined, is an extraordinarily heterogeneous condition. Individual variation is to be expected and, in the current study, is precisely what was observed. Some of the autistic participants read homographs at close to ceiling levels, using the semantic context to select the correct pronunciation of the homograph. Others performed poorly, reading the homographs in a way that was acceptable phonologically but was not always appropriate to the semantic context.

These observations are entirely consistent with our re-analysis of data from previous studies conducted in English (Brock & Caruana, 2014). However, by conducting the study in Hebrew, we were able to test a larger number of homographs. Our mixed random effects analyses indicated that the individual variation in homograph reading was reliable and was not trivially explained away in terms of age differences within the sample. Nor could the variation be fully explained in terms of autism severity, as indexed by CARS score or performance on tests of single word reading accuracy and knowledge of word meaning. These measures all left significant individual variation unaccounted for and none accounted for variation beyond that explained by chronological age. This is not to say that decoding and knowledge of word meaning are unimportant. Clearly, a child with minimal literacy and language comprehension skills would necessarily perform poorly on the homograph test. But within our sample, these were not the main drivers of variation. Instead, somewhat surprisingly, we found that children’s picture naming ability was by far the strongest predictor of homograph reading, accounting for individual variation in performance beyond each of the other predictors and leaving no reliable variation unaccounted for.

There are a number of important limitations of the study to note. First, the sample is small, particularly for analyses comparing the relative strengths of associations between multiple measures. It is also important to acknowledge both the exploratory nature of our analyses and the fact that conclusions rest to some extent on the sensitivity and reliability of each of the predictor measures. Clearly, any conclusions should be considered preliminary, pending a replication with a much larger sample. Nonetheless, the study serves as a demonstration of the potential of our approach to investigating the heterogeneity within the autism population. In particular, our mixed random effects analyses
enabled us to differentiate between genuine unexplained variation in ability and psychometric noise.

While acknowledging the tentative nature of our results, the strong association between homograph reading and picture naming also prompts us to reconsider exactly what the homograph task is measuring and, ultimately, the nature of the impairment affecting those individuals on the autism spectrum who find the task difficult. The conventional view is that homograph reading is a test of language comprehension—performance depends on understanding the meaning of the sentence. Thus, when autistic individuals produce the homograph incorrectly, it is attributed to them failing to understand the sentence as a whole and so processing the homograph out of context. However, studies using other paradigms, including semantic priming, eye-tracking, or reading time measures (Brock, Norbury, Einav, & Nation, 2008; Henderson, Clarke, & Snowling, 2011; López & Leekam, 2003; Norbury, 2005; Saldana & Frith, 2007) have consistently found that autistic individuals exhibit normal context effects in sentence comprehension, at least when overall language ability is controlled for. Here we suggest an alternative explanation—that homograph reading difficulties in fact reflect impairments in the mechanisms of word production rather than comprehension. More precisely, we hypothesize that poor performance on the task originates specifically in a failure to use meaning to select the correct phonological form of the homograph to produce.

Figure 2(a) articulates this hypothesis more formally in terms of contemporary models of reading aloud (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Ellis & Young, 1996; Friedmann & Coltheart, in press). The entry of the homograph in the orthographic lexicon activates two separate entries in the phonological output lexicon corresponding to the two different pronunciations. Selection between these two pronunciations relies on information that comes from the conceptual-semantic system, which determines the relevant meaning of the homograph in the semantic lexicon according to the sentence context. In terms of this model, our hypothesis is that a “disconnection” between the semantic and phonological lexicons prevents the selection of the relevant pronunciation of the homograph.

Figure 2(b) describes the processes involved in picture naming in terms of contemporary cognitive models of word production (e.g., Butterworth, 1989, 1992; Friedmann, Biran, & Dotan, 2013; Garrett, 1992; Levetel, Roelofs, & Meyer, 1999; Nickels, 1997; Nickels & Howard, 2000). Here, the identification of the object in the picture activates the lexical semantic entry, which ultimately guides the selection of the relevant phonological form for output. Crucially, the hypothesized disconnection between the semantic and
phonological lexicons will impair picture naming as well as homograph reading, thereby accounting for the association between performance on the two tasks.

Figure 2 indicates several other points of overlap in the cognitive models of homograph reading and picture naming and, therefore, several alternative accounts of the shared variation in task performance. However, these are less consistent with the available evidence. One such account is that poor performance on both tasks is indicative of impaired access to or retrieval from the semantic lexicon itself, but this fails to explain why picture naming accounted for significant variation beyond that accounted for by word association, our measure of lexical-semantic ability. A related account according to which poor performance in homograph reading and picture naming results from poor vocabulary (i.e., word knowledge) fails on the same grounds. It also struggles to explain why, in previous studies, participants with autism have tended to perform worse on homograph reading than control participants who were matched on their receptive vocabulary knowledge (Happeé, 1997; Joliffe & Baron-Cohen, 1999; López & Leekam, 2003).

Another alternative account is that difficulties occur at the output stage (i.e., within the phonological lexicon or the phonological output buffer). However, this fails to explain why homograph reading was more strongly associated with picture naming than with reading of single words, which also relies on the phonological lexicon. It is also inconsistent with our error analysis (Table 1) which shows that errors were primarily semantic rather than phonological in nature.

An obvious next step would be an attempted replication of our findings in a larger sample of autistic individuals, either conducted in Hebrew or some other language with large numbers of heterophonic homographs. However, our lexical selection hypothesis could also be tested more directly (and in any language) using spoken word production paradigms borrowed from the extensive literature on lexical retrieval deficits in other populations (e.g., Howard & Gatehouse, 2006; Howard, Nickels, Coltheart, & Cole-Virtue, 2006). Analysis of accuracy, error types, and speech production latencies would help elucidate the exact functional locus of impairment in the lexical retrieval process.

Our proposal is a working hypothesis. Irrespective of its ultimate fate, it serves a purpose in focusing research on language production difficulties in autism. Delay in the acquisition of spoken language is common among autistic children, with around one-third remaining “minimally verbal” well into late childhood (Tager-Flusberg & Kasari, 2013). Yet autism has traditionally been characterized as a disorder of language comprehension and semantics (Boucher, 2012; Rutter, 1968). Clearly, language production deficits are not universal in autism, but this does not make them any less important to study. Indeed, as the current study illustrates, important insights may be gained by acknowledging the heterogeneity within the autistic population and harnessing that variability to better understand the linguistic strengths and weaknesses of autistic individuals.

Acknowledgements
Data were presented at the 2012 Icare4autism conference in Jerusalem, Israel and at the 2013 International Meeting for Autism Research in San Sebastian, Spain. We thank Saskia Kohnen and Britta Biedermann for discussion and comments.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Australian Research Council (CE110001021, DP098466), the Israel Science Foundation (1066/14, Friedmann), and a Macquarie University Visiting Research Fellowship.

Notes
1. An exception is the study by Snowling and Frith (1986), which found little difference in performance between children with ASD and intellectually disabled control children of similar language ability. As Happeé (1997) pointed out, in that study, participants were explicitly warned as to the ambiguous nature of the words. Her suggestion, therefore, was that, while they may be able to take account of context when explicitly prompted to do so, the default setting for individuals with ASD is to process words in isolation.
2. Some studies include a fifth homograph, “read,” which involves different tenses of the same verb and, hence, is affected by syntactic context. Frith and Snowling’s (1983) study also included the noun / adjective homograph, “minute” but this was not used in subsequent studies.
3. In Israel, placements in an ASD-specific class is highly regulated, requiring a diagnosis from a child psychiatrist or clinical psychologist, as well as the agreement of a seven member multi-disciplinary panel including a pediatrician, an educational psychologist, a social worker who specializes in children with special needs, the chief inspector of special education in the ministry of education, an inspector of regular education, and a representative of the municipal education committee. We can, therefore, place a high degree of confidence in the diagnoses of these children.

References
American Psychiatric Association (2000). Diagnostic and statistical manual of mental disorders (4th ed., Text Revision). Washington, DC: Author.
American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Publishing.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

Biran, M., & Friedmann, N. (2005). From phonological paraphasias to the structure of the phonological output lexicon. *Language and Cognitive Processes, 20*, 589–616.

Biran, M., & Friedmann, N. (2007). *MA KASHUR: Written and picture association test*. Tel Aviv: Tel Aviv University.

Boucher, J. (2012). Research review: Structural language in autistic spectrum disorder – Characteristics and causes. *Journal of Child Psychology and Psychiatry, 53*, 219–233.

Brock, J. (2011). Complementary approaches to the developmental cognitive neuroscience of autism – Reflections on Pelphrey et al. (2011). *Journal of Child Psychology and Psychiatry, 52*, 645–646.

Brock, J., & Bzishvili, S. (2013). Deconstructing Frith and Snowling’s homograph-reading task: Implications for autism spectrum disorder. *Quarterly Journal of Experimental Psychology, 66*, 1764–1773.

Brock, J., & Caruana, N. (2014). Reading for sound and reading for meaning in autism: Frith and Snowling (1983) revisited. In J. Arciuli & J. Brock (Eds.), *Communication in Autism: Trends in Language Acquisition Research (TILAR) Series* (pp. 125–145). Amsterdam: John Benjamin.

Brock, J., Norbury, C. F., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition, 108*, 896–904.

Burnette, C. P., Mundy, P. C., Meyer, J. A., Sutton, S. K., Vaughan, A. E., & Charuk, D. (2005). Weak central coherence and its relations to theory of mind and anxiety in autism. *Journal of Autism and Developmental Disorders, 35*, 63–73.

Butterworth, B. (1989). Lexical access in speech production. In W. Marslen-Wilson (Ed.), *Lexical representation and process*. Cambridge, MA: MIT Press.

Butterworth, B. (1992). Disorders of phonological encoding. *Cognition, 42*, 261–286.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Ellis, A. W., & Young, A. W. (1996). Human cognitive neuropsychology. Hove, UK: Erlbaum.

Friedmann, N., Biran, M., & Dotan, D. (2013). Lexical retrieval and breakdown in aphasia and developmental language impairment. In C. Boeckx & K. K. Grohmann (Eds.), *The Cambridge handbook of biolinguistics* (pp. 350–374). Cambridge, UK: Cambridge University Press.

Friedmann, N., & Coltheart, M. (in press). Types of developmental dyslexia. In A. Bar-On & D. Ravid (Eds.), *Handbook of communication disorders: Theoretical, empirical, and applied linguistics perspectives*. Berlin/Boston: De Gruyter Mouton.

Friedmann, N., & Gvion, A. (2003). *TILTAN: A test battery for dyslexia*. Tel Aviv: Tel Aviv University.

Frith, U. (1989). *Autism: Explaining the enigma*. Oxford: Blackwell.

Frith, U., & Snowling, M. (1983). Reading for meaning and reading for sound in autistic and dyslexic children. *Journal of Developmental Psychology, 1*, 329–342.

Garrett, M. F. (1992). Disorders of lexical selection. *Cognition, 42*, 143–180.

Geschwind, D. H., & Levitt, P. (2007). Autism spectrum disorders: Developmental disconnection syndromes. *Current Opinion in Neurobiology, 17*, 103–111.

Georgiades, S., Szatmari, P., & Boyle, M. (2013). Importance of studying heterogeneity in autism. *Neuropsychiatry, 3*, 123–125.

Happé, F. G. E. (1997). Central coherence and theory of mind in autism: Reading homographs in context. *British Journal of Developmental Psychology, 15*, 1–12.

Henderson, L. M., Clarke, P. J., & Snowling, M. J. (2011). Accessing and selecting word meaning in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 52*, 964–973.

Howard, D., & Gatehouse, C. (2006). Distinguishing semantic and lexical word retrieval deficits in people with aphasia. * Aphasiology, 20*, 921–950.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition, 100*, 464–482.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Jolliffe, T., & Baron-Cohen, S. (1999). A test of central coherence theory: Linguistic processing in high-functioning adults with autism or Asperger’s syndrome: Is local coherence impaired? *Cognition, 71*, 149–185.

Kuder, G. F., & Richardson, M. W. (1937). *The theory of mental and physical tests*. Hove, UK: Erlbaum.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.

Linzen, T. (2009). *Corpus of blog postings collected from the Israblog website*. Tel Aviv: Tel Aviv University.

López, B., & Leekam, S. R. (2003). Do children with autism fail to process information in context? *Journal of Child Psychology and Psychiatry, 44*, 285–300.

Nickels, L. (1997). *Spoken word production and its breakdown in aphasia*. Hove: Psychology Press.

Nickels, L., & Howard, D. (2008). When the words won’t come: Relating impairments and models of spoken word production. In L. Wheeldon (Ed.), *Aspects of language production*. Hove, UK: Psychology Press.

Norbury, C., & Nation, K. (2011). Understanding variability in reading comprehension in adolescents with autism spectrum disorders: Interactions with language status
and decoding skill. *Scientific Studies of Reading, 15*, 191–210.

Norbury, C. F. (2005). Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders. *Journal of Experimental Child Psychology, 90*, 142–171.

Rutter, M. (1968). Concepts of autism: A review of research. *Journal of Child Psychology and Psychiatry, 9*, 1–25.

Saldana, D., & Frith, U. (2007). Do readers with autism make bridging inferences from world knowledge? *Journal of Experimental Child Psychology, 96*, 310–319.

Schopler, E., Reichler, R. J., & Rochen Renner, B. (1988). *CARS (the Childhood Autism Rating Scale).* Los Angeles: Western Psychological Services.

Snowling, M., & Frith, U. (1986). Comprehension in “hyperlexic” readers. *Journal of Experimental Child Psychology, 42*, 392–415.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: the neglected end of the spectrum. *Autism Research, 6*, 468–478.

Waterhouse, L. (2012). *Rethinking autism: Variation and complexity.* San Diego, CA: Academic Press.

### Appendix 1

**Table 4.** Homographs used in the homograph reading task.

| Homograph | Orthographic transliteration | Pronunciation 1 | Pronunciation 2 |
|-----------|-------------------------------|-----------------|-----------------|
| מְטַר | MTR | matar | Rain | meter | Meter |
| עָז | XOM | xum | Brown | xom | Heat |
| הָעְפַר | HSFR | ha-sapar | The-barber | ha-sefer | The-book |
| כֹּפֶר | KOPH | kofa | Female monkey | kupa | Cashier |
| הָשִׁלְט | HSLT | ha-shelet | The-sign | ha-shalat | The-remote control |
| הָבִרָה | HBIRA | ha-Blra | The-beer | ha-biRA | Capital |
| פרֶס | PRS | peres | Peres | pras | Prize |
| יָנְשֵׁף | INSOF | yanshuf | Owl | yinshof | Exhale-future |
| עֲנֶה | ONA | one | Replies | ona | Season |
| בֵּלֶם | ZKN | zakan | Beard | zaken | Elderly |
| זָרָה | ZR | zer | Crown | zar | Stranger |
| חָזָק | XZKOT | xazakot | Strong-fem-plural | xezkot | Powers |
| סִיני | SINI | sini | Chinese | sinay | Sinay dessert |

Note: Capital letters in the phonemic transliteration indicate the stressed syllable.