

What is 'word understanding' for the parent of a one-year-old? Matching the
difficulty of a lexical comprehension task to parental CDI report.

Suzy Styles and Kim Plunkett
Department of Experimental Psychology
University of Oxford

Abstract

Is parental report of comprehension valid for individual words? If so, how well must an infant know a word before their parents will report it as 'understood'? We report an experiment in which parental report predicts infant performance in a referent identification task at 1;6. Unlike in previous research of this kind (i.e., Houston-Price, Mather & Sakkalou, 2007), infants saw items only once, and image pairs were taxonomic sisters. The match between parental report and infant behaviour provides evidence of the item-level accuracy of both measures of lexical comprehension, and informs our understanding of how British parents interpret standardised Communicative Development Inventories (CDIs).

This is a pre-press proof, reformatted post-publication. Citation: Styles, S J & Plunkett, K (2009) *Journal of Child Language*, 36(4):895-908. doi: 10.1017/S0305000908009264

Introduction

Infant language researchers frequently ask parents to indicate, by means of written comprehension surveys, whether their infants 'understand' various words ranging from 'mouse' to 'yellow' and even 'love'. The question is deceptively simple, yet our assumptions about what we mean by 'understand' or 'comprehend' remain unclear. As one parent jokingly asked, 'Do any of us truly understand *love*?' How high do we, as researchers, set the bar for infant lexical comprehension? And do our participants' parents share the same interpretation of the question?

Vocabulary inventories such as the MacArthur-Bates Communicative Development Inventory (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994) provide a valuable insight into infants' lexical development by tapping into parents' knowledge about their child's daily linguistic interactions. The value of the communicative development inventory (CDI) lies in its ability to assess large sample sizes and large numbers of words more efficiently than would be possible in clinical testing environments, or in labour-intensive journal keeping. Norms for English vocabulary development have been established (i.e., Fenson et al., 1994; Hamilton, Plunkett & Schafer, 2000) allowing an individual's CDI score to be compared with peers of an

appropriate age. CDI scores have also been validated against performance in standardised language tests (Dale, Bates, Reznick & Morisset, 1989; Fenson et al., 1994), making the CDI a valuable tool for assessing infants' general linguistic development.

CDIs asking parents to report on their child's comprehension (i.e., MacArthur CDI, Infant Words and Gestures: Fenson et al., 1994; British CDI: Hamilton et al., 2000) attract particular attention in the literature, where one issue stands out in the debate: While parents can observe concrete instances of lexical production, they can only infer lexical comprehension on the basis of behavioural responses to language (Stiles, 1994).

The problem of inference is compounded by the protracted nature of learning individual words. In learning a new word, first the spoken word must be heard at least once, and be recognised as an isolatable 'chunk' of language (a word). Second, the word must be heard in an appropriate linguistic and social context and in a physical environment where its referent could potentially be distinguished. Third, the word learner must form an association, or 'mapping' between the word and its referent or referential context. Finally, the word learner must check, update and/or refine their initial mapping of the word in order to settle on the same

meaning as other members of their speech community. These elements will require variable amounts of time depending on the frequency at which a particular word is encountered, at what ages, and in what contexts.

Lexical development studies demonstrate that the final stage, gradual acquisition of adult-like meanings, can be particularly protracted. Consistent with Rosch's prototype model of lexical representation (Rosch, 1975; Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976), infants recognise typical referents of words at earlier ages than atypical referents – being able to identify, for example, pictures of robins and labradors as exemplars of 'bird' and 'dog' six months earlier than accepting pictures of ostriches and chihuahuas (Meints, Plunkett & Harris, 1999). Diary studies of infants' early word use also show gradual meaning refinement in early word production, with numerous reports of over- and under-extension during the early stages of acquiring individual words (for overview see Barrett, 1995). It is therefore difficult to identify a discrete cut-off at which a word is unambiguously 'understood' or 'comprehended,' especially given that developments in comprehension can continue even after a word is reliably produced. This presents a problem for parents during CDI assessment, as they are typically given no specific instructions about what level of comprehension is intended for a word to be reported as 'understood' (Tomasello & Mervis, 1994).

So can parents correctly infer which words are 'understood'? And if so, what comprehension cut-off do they use in their judgement? What is needed is external evidence of whether words reported as 'understood' elicit behaviour consistent with lexical comprehension, while words reported as 'not understood' fail to produce this behaviour. Ideally, this evidence should be obtained in a controlled environment where contextual cues (such as daily routine), and social cues (including joint attention and body language) cannot be used to disambiguate the meaning of a word. The inter-modal preferential looking (IPL) paradigm introduced by Golinkoff, Hirsh-Pasek, Cauley and Gordon (1987) offers a possible platform for investigations of this nature. In this paradigm, while two images are presented side by side, the name of one image is presented from a centrally

located loudspeaker above the screen. Looking behaviour can be monitored before and after naming. If naming induces a change in looking behaviour, and leads to a systematic preference for the named image, this can be taken as evidence of lexical comprehension. To date, however, there has been little systematic investigation of IPL performance at the item level, or whether parental report and IPL agree.

Houston-Price, Mather & Sakkalou (2007) report two IPL experiments in which parental CDI report did not predict infants' performance in identifying a named 'target' image. In Houston-Price et al.'s experiments, parents of infants at 1;3, 1;6, and 1;9 filled out two questionnaires, a vocabulary inventory and an object familiarity survey, in the fortnight preceding an experimental session. Experimental stimuli were then tailored to individual infants' lexicons. Items were selected for test if they had been judged as 'familiar' to the infants (encountered more than once a week), and from these items equal numbers of name-known image pairs and name-unknown image pairs were created. Pairs were presented two or four times, with each image acting equally often as a named target, and as an unnamed distracter. Their results showed an increase in target looking following the onset of naming, a finding consistent with lexical comprehension. However, they found no difference between responses to words that were pre-judged to be name-known or name-unknown.

The authors give three possible reasons for their finding. The first two concern inherent parental inaccuracy in CDI reports of comprehension (a failure to observe overt instances of comprehension, such as pointing, or an inability to monitor subtle signs of comprehension, such as eye-gaze). The third notes a possible mismatch in the threshold of 'comprehension' assessed in the laboratory task and in parental judgement. Given the difficulty of establishing a discrete 'cut-off' for comprehension, the current paper explores this third possibility in more detail.

To give an example of mismatch between CDI threshold and IPL task, an infant who is familiar with socks might have a weak representation of the word, which only includes contextual information; associating the word with an indoor scenario, with a particular room in the house, with getting dressed, or perhaps even with

feet. Prior to test, her parents may have judged 'sock' to be 'unknown', as she cannot ordinarily pick out a sock from her trousers. And yet, given the choice between a sock and a lorry, this infant may be able to use her contextual knowledge to infer the correct referent. That is to say, knowing that 'sock' is a household word may be sufficient to correctly infer that the sock is a more likely referent than the lorry. Such a case would represent a mismatch between the level of word knowledge assessed by the parent (sock-trousers), and the difficulty of the particular task (sock-lorry).

A further possibility for the discrepancy between IPL performance and CDI report lies in the sensitivity of the IPL paradigm. Like CDI scores, IPL measures are typically calculated at a global level – assessing infants' behaviour according to an average across multiple trials. Natural variability in infant behaviour may preclude assessment of comprehension at the level of individual items. It is thus important to establish whether the Houston-Price et al. result is based on inaccuracies inherent to the assessment methods (IPL and/or CDI), or whether their findings demonstrate a mismatch between parent and experimenter expectations of what it means to 'understand' a word. If an IPL task can be found in which words reported as 'understood' systematically differ from those reported as 'not understood', and only those reported as 'understood' elicit behaviour consistent with lexical comprehension, then both questions (accuracy, and mismatch) would be addressed. Converging evidence of this nature would provide external validity for the item level accuracy of both IPL and CDI. In addition, the specifics of the IPL task would inform our understanding of the 'level' at which parents judge a word to be 'understood.'

We report an experiment in which CDI report does indeed predict infant performance in an IPL task. In this experiment, picture pairs are from the same semantic category, making it more difficult for an infant to pick out the referent of a weakly-known word (such as 'sock' in the example above). Each item appears only once, acting either as a target or a distracter. We assess the predictive validity of parental report on both naming (change in looking to target images following the onset of the target word) and on image interest (general interest in target and distracter images). In addition,

we address the relationship between vocabulary size and performance.

Method

Participants

Participants were recruited from a database of parents who had previously expressed an interest in participating in developmental studies. Forty one infants between the ages of 1;5.15 and 1;6.21 were initially tested. Six were excluded (2 for parental failure to complete the lexical pre-test, 1 for parental interference during testing, 1 for fussiness, and 2 for experimenter error), leaving 35 infants for whom both CDI information and testing data were available.

Lexical Pre-test

In the week before visiting the laboratory, primary caregivers of all subjects filled out a British adaptation of the MacArthur-Bates CDI (Fenson et al., 1994). The Oxford CDI (Hamilton et al., 2000) is an extended vocabulary checklist of 416 items, selected and adapted from vocabulary sections of the *MacArthur-Bates CDI: Words and Gestures* and the *MacArthur-Bates CDI: Words & Sentences* (Fenson et al., 1994). It is a measure of both receptive and productive vocabulary, in which parents are asked to mark whether infants 'understand' or 'understand and also say' each word in the CDI at the time of survey. The lexical status of each word was classed as 'known' for items marked as either 'understood' or 'understood and also said', and 'unknown' for unmarked items. Parents brought their completed CDIs with them to their testing session.

Materials

Words used in this experiment were selected according to collated data of 548 previously collected British CDIs (Hamilton et al., 2000). 179 CDIs fell between the ages 1;5.15 and 1;6.15. All words selected for use as stimuli were known by more than 50% of infants at 1;6, and met the criteria of being concrete, basic level object names, with referents which were easily imageable. Visual stimuli were created from high quality digital photographs of animals and objects, and presented on a 5% grey background. Photographs depicted animals and vehicles in canonical profile view and

other objects from the perspective of greatest clarity.

Yoked picture-pairs were made up of two images from the same semantic category (animals, vehicles, food, clothing, tableware, furniture), which shared no attested word association in adult language (Moss & Older, 1996), and did not share a phonological onset or rhyme. Taxonomic sisters were selected to increase task difficulty; these items were assumed to be more confusable for infants with only weak representations of the words' meanings, due to shared visual features (such as faces or surface textures), shared perceptual properties (such as having taste, or making loud engine noises), and shared thematic contexts (such as getting dressed, eating, or in the road). The 12 target-distracter pairs were: horse-lion, cow-chicken, dog-monkey, mouse-bear, banana-toast, cake-bread, bus-lorry, bicycle-pushchair, coat-sock, shoe-pyjamas, cup-bowl, and bed-highchair.

For audio stimuli, a single digital recording session was conducted in a sound-attenuating booth on DAT tape sampling at 44.1 kHz. Three tokens of each auditory stimulus were produced by a female native speaker of British English, using high-affect child directed speech. The single best token of each stimulus was manually selected for clarity, typicality and affect, and edited to remove head and tail clicks, and to align the onset of the target word to 2500ms after the start of the trial.

Procedure

After a few minutes of 'settling in', infants sat on a caregivers' lap facing a rear-projection screen in an IPL booth. Parents were asked to wear headphones and to close their eyes during the procedure, which lasted approximately one and a half minutes. 12 trials were presented. A picture-pair appeared on-screen for 5000ms, during which an auditory phrase began (e.g., 'Look at the dog!'). The target word began 2500ms into the trial, creating a pre- and a post-naming phase of equal length. All infants saw the same 12 picture pairs, in which no item was repeated. Trial order was randomised on presentation, and target side was counterbalanced. Between trials, an auditory distracter phrase was played. This phrase oriented infants' attention to the centrally mounted speaker above the screen, disrupting the likelihood of trial-to-trial side-bias,

or semantic interference. Infants sat approximately 90cm from a screen with a display area 790mm wide. Images were 320mm wide, together occupying a visual angle of 48° and separated by a gap of 155mm (10°).

Scoring

Infants' eye movements were monitored by small cameras located above the two image presentation areas. Recordings were digitally captured as a split-screen image during test. Blind manual coding was conducted offline frame-by-frame at a temporal accuracy of 40ms using Noldus Observer software. All coding was conducted by an experienced coder (previously assessed inter-coder reliability above 0.95). Looks to left and right were automatically re-combined with trial information. Looks to the target and distracter were analysed separately for the pre- and post-naming phases of the trial.

Our primary hypothesis is that if the current IPL task taps into the same level of specificity as parental interpretation of the CDI question, then looking to the target image will increase in the post-naming period only when targets are reported as known, but will not increase when the target is reported as unknown. Knowing the name of the distracter may well influence picture looking, but it is not predicted to interact with the effect of naming a known target.

Analyses

Two measures of target looking were calculated for each trial: The proportion of target looking (PTL) is defined as the total amount of time an infant spends fixating the target (T) as a proportion of the total amount of time she spends fixating target and distracter (T+D), expressed as $T/(T+D)$. Blinks, switches between images, and looks away from image areas are excluded from this measure. The difference in longest look (LLK) is defined as the difference between the single longest look to the target (t) and the single longest look to the distracter (d), expressed as $t-d$. Both measures are calculated separately for the pre-naming and the post-naming period, with the pre-naming period representing baseline image interest for each trial. A systematic increase in target looking following

the onset of ‘naming’ is taken as evidence of lexical comprehension.

Following test, trials were categorized according to parental report of target status (known, unknown) and distracter status (known, unknown), resulting in four trial types: (target known, distracter known; target known, distracter unknown; target unknown, distracter known; target unknown, distracter unknown). As the identification of trial types was a post hoc assessment, there was a possibility of skewed distribution between the trial types. The data were therefore analysed twice: In the broad analysis, data from all 35 participants were included. Pre-tests confirmed a distributional skew in the broad analysis towards knowing the target (fifteen infants, for example, knew more than 10 of the 12 target words). In the narrow analysis, analysis was restricted to those infants whose trials were spread across all four trial types. 16 participants fulfilled this criterion. This paper therefore reports the results of the more stringent, narrow analysis, where the number of targets and distracters known to infants was very closely balanced. The same pattern of results was obtained for both the broad and the narrow analysis.

Results

According to parental CDIs, infants included in the narrow analysis knew a mean of seven of the target words, ranging from three words (known by one infant) to nine words (two infants), and a mean of five distracter words, ranging from two words (one infant), to seven words (two infants). The best known target words were *dog* and *shoe*, known by 15 of 16 subjects. Targets were known by a mean of seven infants (range: 3 to 15). The best-known distracters were *sock* and *pushchair*, known by 11 of the 16 infants. Distracters were also known by a mean of seven infants (range: 2 to 11). Of the sixteen infants’ 192 trials, the target was reported as known in 54% of trials, while the distracter was known in 41% of trials. Across the four trial types, 22% were target known, distracter known, 32% were target known, distracter unknown, 18% were target unknown, distracter known, and 28% were target unknown, distracter unknown. The mean receptive CDI score for infants in the narrow analysis was 113 words of a possible 416 ($SD=53$), ranging from 44 words to 187 words.

Participant means were calculated for each trial type. In order to assess the relative impact of target and distracter status, a three way repeated-measures ANOVA was conducted comparing the effect of naming (pre-naming, post-naming), in conjunction with target status (known, unknown) and distracter status (known, unknown) on the proportion of target looking (PTL). There was a main effect of distracter status ($F(1,60) = 12.54$, $p=.001$, partial $\eta^2=.17$), an interaction between target status and distracter status ($F(1,60) = 5.65$, $p=.02$, partial $\eta^2=.09$), and an interaction between target status and naming ($F(1,60) = 5.82$, $p=.02$, partial $\eta^2=.09$). Simple effects clarified that when the distracter was known, infants looked at targets less than when they did not know the name of the distracter (distracter known: $M=.45$, $SD=.12$; distracter unknown: $M=.55$, $SD=.10$; $t(62) = 3.34$, $p=.001$, $d=0.87$). This was an overall effect, and did not interact with target naming. Concerning the interaction between target status and distracter status, when the target was known, distracter status had no impact on target looking (distracter known: $M=.50$, $SD=.10$; distracter unknown: $M=.53$, $SD=.09$; $t(30) = .93$, n.s.). When the target was

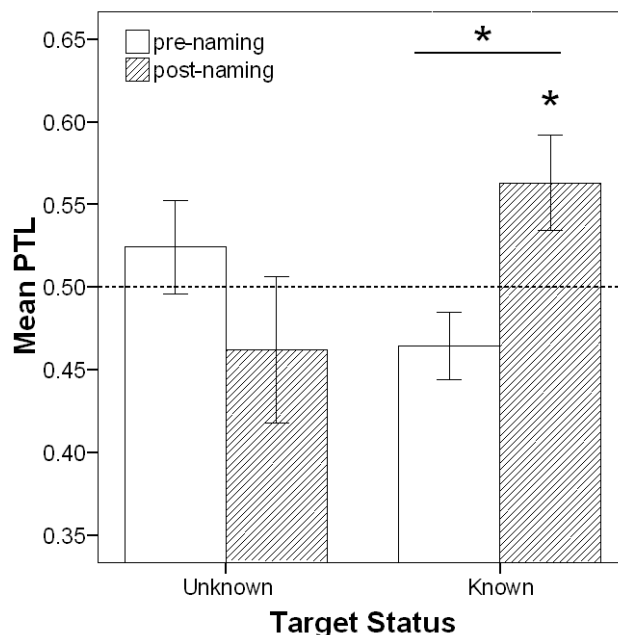


Figure 1. Proportion of target looking pre- and post-naming, according to target status. Mean PTL in narrow analysis (N=16). Error bars show +/- 1 standard error. * indicates $p<0.05$. Target-known trials show significant target preference in the post naming period. $t(31)=2.18$, $p=.04$, $d=.39$.

unknown, known distracters reduced overall target

looking (distracter known: $M=.41$, $SD=.13$; distracter unknown: $M=.57$, $SD=.10$; $t(30) = 3.81$, $p=.001$, $d=1.35$). This was again an overall effect and did not interact with naming (pre- or post-naming).

As the primary investigation concerns whether parental report identifies items which attract behaviour consistent with lexical comprehension, the interaction between target status (known, unknown) and naming (pre-naming, post-naming) is especially informative. In trials where the target was reported as known, there was a systematic increase in target looking following naming (pre-naming: $M=.46$, $SD=.12$; post-naming: $M=.56$, $SD=.16$; $t(31) = 2.63$, $p=.01$, $d=.70$). In trials where the target was reported as unknown, there was no systematic change following naming (pre-naming: $M=.52$, $SD=.16$, post-naming: $M=.46$, $SD=.25$; $t(31) = 1.12$, n.s.). This finding indicates that infants only increase their looking to named targets when their parents have reported the target word to be 'understood'.

Analysis of LLK also produced the same pattern of effects and interactions, and is omitted for brevity. Further support for this pattern of response comes from the broad analysis of data from all 35 subjects, which produced the same patterns of results in both PTL and LLK.

In summary, target-known trials showed a significant increase in target preference following the onset of naming. By contrast, the target-unknown trials showed no systematic change in image preference, with preference for both images remaining similar throughout. Knowing the name of the distracter also affected the looking pattern at the overall level, but did not interfere with the general effect of naming on target recognition. This finding demonstrates a match between an IPL task and CDI report for a particular population of infants (British 18-month-olds), whereby parental report of individual words predicted which words would attract looking behaviour consistent with lexical comprehension.

One additional possibility remains for interpretation of these results. If IPL indexes infants' general linguistic ability, regardless of item, then infants with larger vocabularies may be leading the trend, as the target-known mean for large vocabulary infants would contain more trials than the target-unknown mean (potentially resulting

in a higher likelihood of success). To address this possibility, an analysis was conducted to evaluate the impact of vocabulary size on performance, using the change in PTL following naming as the dependent measure. It was predicted that if change in PTL indexed general task performance, then infants at the high and the low end of the vocabulary scale would perform differently for words reported as 'known', with higher performance for large vocabulary infants. If this were the case, for 'known' words, change in PTL would be expected to co-vary with infants' CDI comprehension score.

When the target word was reported as known, no correlation was evident between each infants' CDI comprehension score and the change in PTL following naming ($\rho(32)=.03$, n.s.). For simple comparison, infants were also separated into two CDI groups according to a median split of their comprehension score. No difference was evident between low- and high-vocabulary infants ($t(30)=1.86$, n.s., $d=0.17$). This finding indicates that for 'known' words, infants across the vocabulary range presented the same pattern of results: an above-chance increase in target looking ($t(31)=2.63$, $p=.01$, $d=0.47$). These findings demonstrate that the general pattern of target discrimination for known words was not driven by general linguistic development, but by accurate parental identification of items attracting performance consistent with lexical comprehension – a finding which is true for both large and small vocabulary 18-month-olds.

Discussion

Our results demonstrate that parental report can predict the group of words infants will be able to identify in an IPL paradigm, when the task has certain characteristics. In our task, infants had only one chance to identify a named target, and to pick it out from another item from the same semantic category. This match between CDI threshold and IPL task provides external validity for the accuracy of both measures of lexical comprehension at the item level: The finding suggests that parents' inferences about their infants' 'understanding' may be more accurate than previously acknowledged; it also suggests that IPL is sensitive to lexical comprehension of individual items, even in 'one-

shot' designs, despite the usually global level of assessment.

These findings differ from the results of Houston-Price et al. (2007), which showed performance consistent with lexical comprehension for all items, whether the target had been rated as known or unknown. The Houston-Price et al., experiment shares many similarities with the current experiment: both experiments assessed lexical comprehension at 1;6 using the same CDI, with the same instruction sheet; the IPL experiments contained trials of the same duration, presented on-screen in the same manner, with target words occurring in similar pre-recorded phrases; infants sat with their parents, who were given the same instructions, and the experimenter was not present during test. Thus the discrepancy in results must lie, not in overall methodological differences, but in the particular stimulus combinations.

Two key differences in stimulus preparation can be highlighted, both of which affect the difficulty of the task: distracter selection, and stimulus repetition. In the current study, all picture-pairs contained a target and a distracter from the same semantic category. In the pair, 'sock-trousers', for example, both images were clothing, made of fabric, likely to be encountered in the routine 'getting dressed,' which typically occurs indoors. Infants therefore needed to know more about target words in order to distinguish them from this type of distracter: Knowing that 'sock' is related to getting dressed would be insufficient to pick out a sock from trousers. An additional level of difficulty in the current experiment was the 'one-shot' design, in which infants encountered each item only once. There are a number of reasons why a single presentation of picture pairs should be predicted to make an IPL task difficult. Firstly, considering that infants have only a brief exposure to the test exemplars prior to naming (2,500ms), and only a short amount of time (2,500ms) to demonstrate comprehension following naming, if an infant needs more time to decide whether a novel exemplar matches a familiar label, then they will not be able to demonstrate comprehension in a 'one shot' experiment. On a second presentation however, the familiarity of the exemplars could reduce the difficulty of recognising a 'match'. Secondly, successful identification of a known

target on the first presentation of a yoked picture pair may facilitate mapping of an unknown word in the second presentation, via mutual exclusivity (Markman, 1989; Merriman & Bowman, 1989) or N3C (novel-name-nameless-category: Golinkoff, Mervis & Hirsch-Pasek, 1994). Such effects may be weaker if the unknown label is presented first, however repetition of picture pairs gives infants a more opportunities to clarify or consolidate a weak representation during testing.

As pointed out by Houston-Price et al., multiple presentations of stimuli, in which an image appears equally often as a named target, and as an unnamed distracter, masks whether infants employ mutual exclusivity or N3C to infer a correct referent. If this were the case, one would expect to see an interaction between distracter status and naming, such that looking to targets increased when the distracter was known and the target was not. An additional advantage of the 'one-shot' design was that it allowed us to independently assess the status of target and distracter words. No such pattern was evident in the current experiment. While distracter status did have an overall effect on target preference in this task, it did not interact with the change caused by naming. This finding fits well with evidence that mutually exclusive responding requires a certain amount of processing time. Mather and Plunkett (In Press) for example, report an IPL experiment in which a novel object was presented alongside a familiar object, while a novel word was heard. Post-naming preference for the novel object did not appear on the first presentation of object pairs, but was observed on subsequent presentations. Their finding is consistent with the current observation that when only the distracter was name-known, infants did not show mutual exclusivity style target preference following naming, when they had only one presentation of pictures.

One final difference in stimulus selection remains. In the Houston-Price et al. task, all stimuli were selected from a pool of items which had been reported to be encountered at least weekly. In the current experiment, stimuli were selected from a pool of words reported as 'understood' by a large cohort of eighteen-month-olds. The cohort comprehension data provided us with an approximation of item familiarity at this age (on the logic that items typically encountered frequently

are more likely to be understood). However, no direct measure of object familiarity was included in the current design. This was, in part, to ensure that all infants were able to view the same set of stimuli. It is therefore possible that object familiarity played some part in infants' different responding to words reported as known and unknown. As item familiarity, item repetition and distracter-type are confounded between the two studies, it is impossible to draw direct conclusions from their comparison. A more systematic investigation of these elements (i.e., a one-shot IPL design pitting object familiarity against distracter-type) would be needed to clarify the exact contribution of each. Despite this limitation, the results of the current experiment still contribute new insights to the kind of IPL task which can match parental CDI reports, and the accuracy with which both IPL and CDI approximate the nebulous concepts of 'comprehension,' and 'understanding.'

Conclusions

In this study we find binary parental assessments of word 'understanding' can predict the trials in which 18-month-old infants will identify named targets in an IPL task. Like CDI scores, IPL behaviours are typically used to produce a global assessment, with performance averaged across a number of trials and a number of infants. This new finding lends support to both the accuracy of CDI report, and the accuracy of IPL at the item level, in-so-far-as the two tools agree on which words an individual understands. We would not want to claim that IPL provides 100% accurate assessments of lexical comprehension for all infants, on all items, all of the time. Nor would we want to claim that CDI reports of comprehension are 100% reliable for all parents, on all items, all of the time. However,

converging evidence from these two sources demonstrates that parental report on-the-whole identifies those items which will attract behaviour consistent with lexical comprehension in a stringent laboratory test, and distinguishes them from items where performance will be unsystematic.

Furthermore, the match between CDI and IPL informs our understanding of the 'level' of parental CDI responses. Given the difficulty of deciding the threshold for when a word is 'known,' 'comprehended,' or 'understood,' establishing the parental threshold is important. For British parents, 'Does your child understand the word 'mouse'?' could become either 'Can your child pick out a mouse from a bear?' or 'Does your child know enough about the word 'mouse' to know that it's not a lorry?' We believe the current experimental task closely matches parents' interpretation of what it means to 'understand a word': British parents of 18-month-olds mark as 'understood' those items which their infants will be able to identify correctly, with only one presentation, in an unfamiliar environment, in the presence of potentially confusing distracters. While it remains to be seen how much item familiarity influences infants' pattern of responding, it appears that British parents set a high bar for their infants' 'understanding' of object names. Future research could establish whether parental judgements differ when different instructions are included with CDIs.

Acknowledgements. This research was supported by a Clarendon Fund Bursary to Suzy Styles, and grants to Kim Plunkett from the ESRC (Grant No. RES-000-23-1322) and the Leverhulme Trust. The authors would like to thank the staff of the John Radcliffe Hospital for their ongoing support.

References

Barrett, M. (1995). Early lexical development. In P. Fletcher & B. MacWhinney (eds.), The Handbook of Child Language. Oxford: Blackwell.
 Dale, P. S., Bates, E., Reznick, J. S. & Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. Journal of Child Language 16, 239-49.
 Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J. & Pethick, S. J. (1994). Variability in

early communicative development. Monographs of the Society for Research in Child Development 59, 1-173.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M. & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. Journal of Child Language 14, 23-45.

Golinkoff, R. M., Mervis, C. B. & Hirsch-Pasek, K. (1994). Early object labels: The case for a

developmental lexicals principles framework.
Journal of Child Language 21, 125-55.

Hamilton, A., Plunkett, K. & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language* 27, 689-705.

Houston-Price, C., Mather, E. & Sakkalou, E. (2007). Discrepancy between parental report of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language* 34, 701-24.

Markman, E. M. (1989). *Categorization and naming in children*. Cambridge: MA: MIT Press.

Mather, E. & Plunkett, K. (In Press). Learning words over time: The role of stimulus repetition in mutual exclusivity. *Infancy*.

Meints, K., Plunkett, K. & Harris, P. (1999). When does an ostrich become a bird? The role of typicality in early word comprehension. *Developmental Psychology* 35, 1072-8.

Merriman, W. E. & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning.

Monographs of the Society for Research in Child Development 54, 1-132.

Moss, H. E. & Older, L. (1996). *Birkbeck Word Association Norms*. Hove, UK: Psychology Press.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104, 192-233.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology* 8, 382-439.

Stiles, J. (1994). On the nature of informant judgements in inventory measures. and so what is it you want to know? *Monographs of the Society for Research in Child Development* 59, 180-5.

Tomasello, M. & Mervis, C. B. (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development* 59, 174-9.